8-9-2019

Electrophysiological, Neural, and Perceptual Aspects of Pitch

Karl D. Lerud
University of Connecticut - Storrs, karl.lerud@uconn.edu

Electrophysiological, Neural, and Perceptual Aspects of Pitch

Karl D. Lerud, PhD

University of Connecticut, 2019

Pitch is a perceptual rather than physical phenomenon, important for spoken language use, musical communication, and other aspects of everyday life. Auditory stimuli can be designed to probe the relationship between perception and physiological responses to pitch-evoking stimuli. One technique for measuring physiological responses to pitch-evoking stimuli is the frequency following response (FFR). The FFR is an electroencephalographic (EEG) response to periodic auditory stimuli. The FFR contains nonlinearities not present in the stimuli, including correlates of the amplitude envelope of the stimulus; however, these nonlinearities remain undercharacterized. The FFR is a composite response reflecting multiple neural and peripheral generators, and their contributions to the scalp-recorded FFR vary in ill-understood ways depending on the electrode montage, stimulus, and imaging technique. The FFR is typically assumed to be generated in the auditory brainstem; there is also evidence both for and against a cortical contribution to the FFR. Here a methodology is used to examine the FFR correlates of pitch and the generators of the FFR to stimuli with different pitches. Stimuli were designed to tease apart biological correlates of pitch and amplitude envelope. FFRs were recorded with 256-electrode EEG nets, in contrast to a typical FFR setup which only contains a single active electrode. Structural MRI scans were obtained for each participant to co-register with the

electrode locations and constrain a source localization algorithm. The results of this localization shed light on the generating mechanisms of the FFR, including both cortical and subcortical sources.

Electrophysiological, Neural, and Perceptual Aspects of Pitch

Karl D. Lerud

B.F.A, University of Wisconsin-Milwaukee, 2006

M.A., University of Wisconsin-Milwaukee, 2011

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy at the University of Connecticut

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Electrophysiological, Neural, and Perceptual Aspects of Pitch

Presented by

Karl D. Lerud, B.F.A., M.A.

Major Advisor

(Edward W. Large)

Associate Advisor

(Erika Skoe)

Associate Advisor

(Heather Read)

Associate Advisor

(Whitney Tabor)

University of Connecticut

2019

Contents

Chapter 1

Nonlinear frequency components in auditory responses to complex sounds

1.1 Introduction

The auditory system is interesting and peculiar in many ways, especially the ways in which it is nonlinear. In particular, auditory physiological responses exhibit properties such as extreme amplification and compression and sharp frequency selectivity (Eguíluz et al., 2000). But one of the most marked and universal characteristics of auditory responses is the generation of additional frequency components in response to complex stimuli. While the various compressive effects are readily observable in the responses to even a single

sinusoid, complex stimuli are generally used to clearly observe the nonlinear consequence of additional response frequencies beyond what is contained in a stimulus. Here we focus on these nonlinear frequency components (NFCs) and their characteristics.

There are three dominant techniques for recording these particular physiological responses from the auditory system, and each will be reviewed and explored here. Far-field potentials are electrical responses generated by the cochlea and brain but recorded with electrodes at some distance from their sources, often from the scalp. Laser velocimetry is a technique in which a reflective bead is placed directly on the basilar membrane and its velocity is recorded. A laser is pointed at the bead and its Doppler shift is recorded from its reflection. This technique is also called laser Doppler velocimetry, laser interferometric velocimetry, or laser interferometry, and has had a wide array of applications in physics and engineering previous to its adoption in auditory neuroscience. Otoacoustic emissions (OAEs) are pressure waves generated mechanically within the cochlea, and are recorded with a microphone inside the ear canal. The portions of these emissions which consist of NFCs are called distortion product otoacoustic emissions (DPOAEs).

Combining knowledge of features and limitations of these techniques, along with some conclusions from invasive animal research, we will attempt to draw conclusions about the generating mechanisms of these NFCs. In the case of far-field potentials, a periodic and roughly phase-locked auditory response to a periodic stimulus is known as a

frequency-following response (FFR) (Worden and Marsh, 1968). There has been much debate in the literature about the location or locations responsible for generating the scalp-recorded FFR (Bidelman, 2015; Chandrasekaran and Kraus, 2010; Gardi et al., 1979). The FFR often contains the NFCs we will be interested in, so information about these sites of generation may provide insights into the mechanisms giving rise to the NFCs themselves. In the case of otoacoustic emissions, it is quite clear that the site of generation is within the cochlea, because it is a pressure wave. However there is ample evidence that DPOAEs are modulated in their amplitude and phase by contralateral stimulation through efferent circuits, mostly the medial olivocochlear bundle (Abel et al., 2009; Deeter et al., 2009; Kujawa, Fallon, et al., 1995; Kujawa and Liberman, 2001). This suggests that the mechanism of generation of these NFCs may not be found solely within the ipsilateral cochlea. In the case of laser velocimetry, the situation is similar. The basilar membrane is the source of the movement, but the NFC portion of the response may have roots in places other than the basilar membrane itself. Through converging evidence from these three methods, we will see that NFCs are likely not entirely mechanical, but also neural in origin.

The simplest type of complex stimulus is one which contains precisely two sinusoids of different frequencies, thus this kind of stimulus has been used frequently in the literature exploring the auditory system at various levels. There are two dominant NFCs that will chiefly interest us here: the cubic difference tone (CDT) and the quadratic difference tone

(QDT). In the case of a two-frequency stimulus composed of sinusoids at f_1 and f_2 with $f_1 < f_2$, the frequency of the CDT is $2f_1 - f_2$ and that of the QDT is $f_2 - f_1$. Thus if $f_1 = 400$ Hz and $f_2 = 500$ Hz, the CDT is 300 Hz and the QDT is 100 Hz. The QDT is a second-order nonlinearity and the CDT is third-order. In general there are two types of responses to complex sounds: Even-order and odd-order. Though not always the most prominent components, responses to the stimulus primaries themselves are odd-order responses, namely first-order. The formula for the second-order QDT can generalize to all even-order NFCs of $kf_2 - mf_1$ Hz, where $k > 0$ and $m = k$ or $m = -k$. Notice that even this simple formula predicts that even-order NFCs are not only lower than f_1 but also higher than f_2, such as the summation tone $f_2 + f_1$. The formula for the third-order CDT can generalize to all odd-order NFCs of $kf_1 - mf_2$ Hz, where $k - m = 1$. Here again we see that this also predicts NFCs higher than f_2, in the case that $\{k, m\} < 0$. All of these NFCs can be recorded to some extent in the auditory system, and we will see in what situations higher-order NFCs are readily observed, for instance in laser velocimetry studies with certain combinations of f_2/f_1 ratio and value of f_1 (e.g. Robles et al. (1997)).

There is a general way to separate odd-order from even-order responses. Auditory responses are often evoked, meaning a large number of short trials are done and the responses to them are averaged. However, many researchers deliver half of their stimuli in one polarity, and the other half in the opposite polarity. This is equivalent to flipping the

stimulus over, or multiplying it by -1, for half of the presentations. Even-order responses do not alternate polarity with the stimulus, but odd-order responses do, therefore if the two groups of responses were stored separately, one can average them separately, and then add the averages together to isolate the even-order responses, and subtract the averages to isolate the odd-order responses (Lerud et al., 2014, Appendix A). Historically, researchers use a version of this approach to avoid stimulus transducer electromagnetic contamination in the recording mechanisms (Skoe and Kraus, 2010): They simply deliver the stimuli in alternating polarity, but then do not store the respective responses separately, instead just averaging them all together at the end. It is apparent that this is the same thing as the procedure to isolate the even-order responses described above, differing only by a factor of 2, while throwing away the odd-order responses. If stimulus artifact is contained in auditory responses, because it is first-order, it will alternate polarity with the stimulus presentations, and will thus not be present in the evoked response obtained in this way. However, it is important that researchers realize that responses obtained in this way have thrown away not just any potential stimulus artifact, but also half of the actual auditory response, namely the odd-order portions of it. This includes, as described above, responses to the primaries because these are also first-order responses, any third-order nonlinearities such as the CDT, any quintic nonlinearities, and so on. These issues come up most prominently in the FFR literature, and will be further discussed below.

1.2 Far-field potentials

Of the three to be discussed, the far-field potential is the oldest technique. As amplification and other hardware has improved in the past decades, modern recordings of this type are typically taken from electrodes on the scalp, and can therefore be obtained non-invasively either from humans or other animals. However the technique developed from multi-unit recordings taken from gross electrodes positioned within axonal bundles or nuclei. Responses obtained in this manner are qualitatively similar to potentials recorded from the scalp, because in both cases the post-synaptic potentials along a very large number of neurons need to sum to a well-defined signal in order for any response to be observed. If each of the neurons being recorded were encoding an auditory stimulus in its own unique way, population responses of this type would likely not contain significant signal.

Some of the first auditory responses recorded in this way were those of Wever and Bray (1930a). They used a copper electrode on and around the auditory nerve and observed a response mirroring the sinusoidal stimuli they were delivering. This was attributed to population-level neural activity within the auditory nerve in both Wever and Bray (1930a) and Wever and Bray (1930b). However it was subsequently found that they were in fact observing a non-neural response. The activity they recorded from their electrode was exactly in phase with the stimulus itself, which would only be possible if it were being generated immediately upon the stimulus altering the pressure in the cochlear

fluid. This mechanical action is what causes a portion of the basilar membrane to resonate concurrently with the pressure wave in the fluid, and this in turn depolarizes the hair cells. This depolarization, despite not being neural, generates a significant enough electrical response to be recorded at large distances from the cochlea, and was easily picked up by Wever and Bray's electrode. This response came to be known as the Wever-Bray response, and, more commonly today, the cochlear microphonic (CM). These and many other relevant facts are pointed out by Worden and Marsh (1968), which was the first significant characterization of population-level auditory responses of this type, and was the first work to define the frequency-following response (FFR).

We will use FFR to refer to far-field potentials, whether they come from invasive gross electrodes or scalp electrodes, because we will only be considering periodic responses to periodic stimuli, and not for instance transient-evoked or spontaneous auditory responses. Worden and Marsh (1968) worked with cats, measuring FFRs to single sinusoids from many locations in the auditory pathway, from the cochlear nuclei up to various areas of cortex. The last place they observed an FFR was the inferior colliculus; nothing resembling an FFR occurred in cortex. This finding is largely maintained to the present, although as both hardware and software have become more advanced, there has been some recent evidence of cortical contributions to the FFR, e.g. Coffey et al. (2016).

Worden and Marsh (1968) also pointed out several ways in which a researcher may

tell whether a response is neural or cochlear in origin. One way is to observe whether or not the response is delayed. Their most robust recordings come from the cochlear nucleus, and they noted, as stated above, that neural responses are delayed in phase with respect to the stimulus. The reason for this is that the signal has already passed through at least one synapse (auditory nerve dendrites on the hair cells) or more (auditory nerve afferents on the cochlear nucleus). This delay is typically between 2 and 3 milliseconds (Greenberg et al., 1987), and is readily visible in their plots. Another way to tell whether the response is cochlear or neural is whether it is thresholded. They observe that the neural responses appear thresholded, that is, there is a very rapid onset of the FFR when the stimulus reaches a certain amplitude, even if the stimulus is ramped on. By contrast, the CM grows as an approximately linear function of the stimulus, and is of course also in phase with it. The authors measured the CM separately from the FFR in this paper by recording from an electrode placed on the round window of the cochlea, a place reasonably sure to record a strong CM but minimal neural response.

With these aspects of the FFR in mind, the scalp-recorded far-field potential becomes more complicated and interesting. It is apparent that, without suitable control methods, the scalp-recorded FFR will contain responses from a wide variety of sources in the subcortical auditory system, including the cochlea itself. Moushegian et al. (1973) were the first to document a scalp-recorded FFR in humans, taking influence from the growing body

of research in animals suggesting that the FFR was a prominent and important part of the auditory system's operation. They took pains to assure the reader that what they recorded was strictly neural, and not the CM or a simple stimulus artifact, the most relevant piece of evidence being the latency of the response. While in Worden and Marsh (1968) they saw a roughly 2 millisecond delay from the cochlear nucleus, here the authors observed a 6 millisecond delay. This is good evidence that the scalp-recorded FFR is being generated in the upper brainstem, likely either the lateral lemniscus or inferior colliculus.

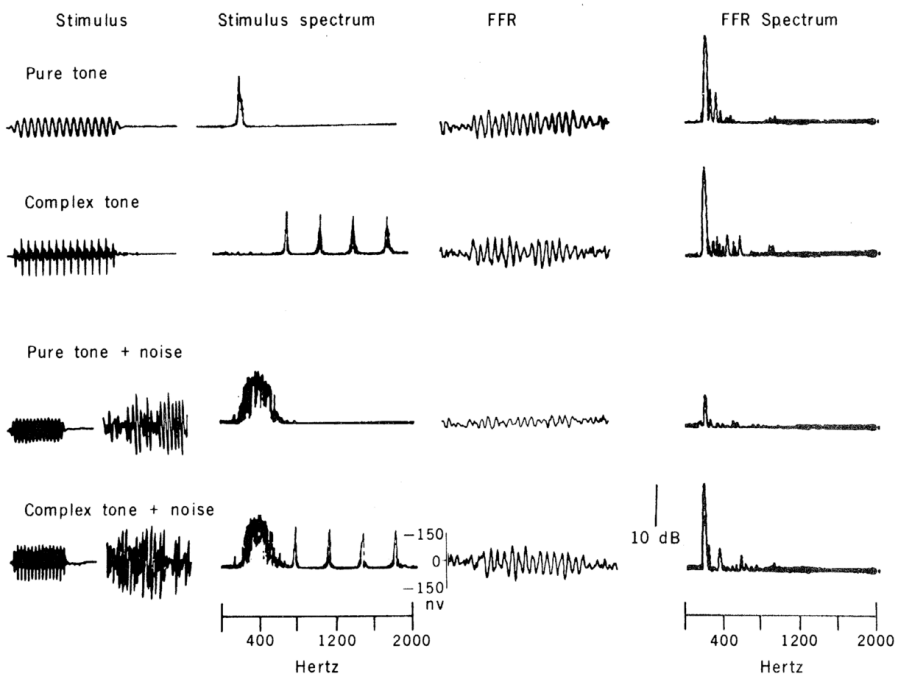


Figure 1.1: Summary plot of results from Smith et al. (1978).

All stimuli in the above studies were pure sinusoids, so there was no opportunity to observe NFCs. Studies around this time period suggested that "combination tones" (what we are calling NFCs) were characteristics of auditory responses through the autocorrelation of single-unit recordings in the auditory nerve (J. L. Goldstein and Kiang, 1968), but researchers had not yet tried multicomponent stimuli in human FFR experiments. The first such effort was Smith et al. (1978), who investigated "missing fundamental" stimuli. It has been noted for a long period that the pitch of a harmonic, complex tone with no spectral energy at the fundamental frequency is nevertheless perceived to be the fundamental. These researchers designed a simple FFR experiment to look for neural correlates of the pitch percept at the fundamental. The complex tone the researchers used contained harmonics 2 through 5 of the missing fundamental. Thus we can note, with hindsight, that they did not distinguish in this experiment between the CDT and the QDT. We recall that the CDT is $2f_1 - f_2$ and the QDT is $f_2 - f_1$. Taking the second harmonic as f_1 and the third harmonic as f_2 in this situation, it is clear that the CDT and QDT are the same frequency, namely the missing fundamental, since $2 * 2 - 3 = 3 - 2 = 1$. They in fact showed a robust response at this frequency, which was 365 Hz (Figure 1.1).

They also made an additional contribution to the question of the neural origin of the FFR, and of the NFC in particular. In the complex stimulus, they also added a narrowband noise in the region around the fundamental. If the cochlea were generating the

NFC as a result of a pattern of excitation in the region of the basilar membrane resonant with the NFC, then the noise should interrupt and cancel that resonance to a significant extent. As a result of this, the NFC should not be readily visible in the scalp-recorded FFR, or should at least be significantly attenuated. These researchers found that the NFC was hardly attenuated for the complex tone: only 1 dB less than the condition without noise. This basic finding has since been replicated by Smalt et al. (2012) who used complex tones consisting of much higher harmonics. They confirmed that the QDT at the fundamental was preserved in the presence of lowpass noise below the stimulus primaries, whereas the CDT and other higher, odd-order NFCs were significantly attenuated in the presence of the noise.

The first study to actively seek out and characterize NFCs in the FFR was Chertoff and Hecox (1990). This study sought to look at NFCs from a clinical and diagnostic standpoint, with the expectation that the auditory system in healthy people should be reliably nonlinear in consistent ways. They observed extensive evidence of the QDT and its harmonics in both guinea pigs and healthy humans for multiple pairs of stimulus primaries, and also show an almost complete lack of any FFR in two deaf humans. This is also the first FFR study to show how to selectively gather either even-order or odd-order NFCs as described above, and they correctly characterize the reason why they only saw even-order NFCs, which is that they used stimuli with alternating polarities and averaged the

responses to the two polarity conditions. Proper electromagnetic shielding is always desirable, but in case there is any worry that the transducer will induce a current directly into the electrodes, also called stimulus artifact, some researchers opt for the polarity reversal technique. If there is any stimulus artifact present in the electrode recordings, it will shift its polarity along with the actual stimulus and be canceled out in the subsequent average.

Wondering why they did not record any evidence of the CDT having recorded their FFRs in this manner, Chertoff and Hecox (1990) correctly state the following: "the odd-order distortion products were canceled by averaging responses to signals that alternated in polarity. Odd-order distortion products, such as the CDT, follow the polarity of the stimulus; that is, if the polarity of the response is positive when the stimulus is positive and negative when the stimulus is negative, then averaging the responses would eliminate any evidence of the CDT." They also lowpass filtered their responses below the stimulus primary frequencies, to eliminate stimulus artifact. Yet they were already doing that by alternating the polarity of the stimuli and averaging the responses. If they hadn't lowpass filtered, they would have found no response at any of the primary frequencies, and this would be entirely consistent with the quoted statement above, since these are in fact odd-order responses, namely first-order.

These lines of thought are continued in Rickman et al. (1991), a bigger study with

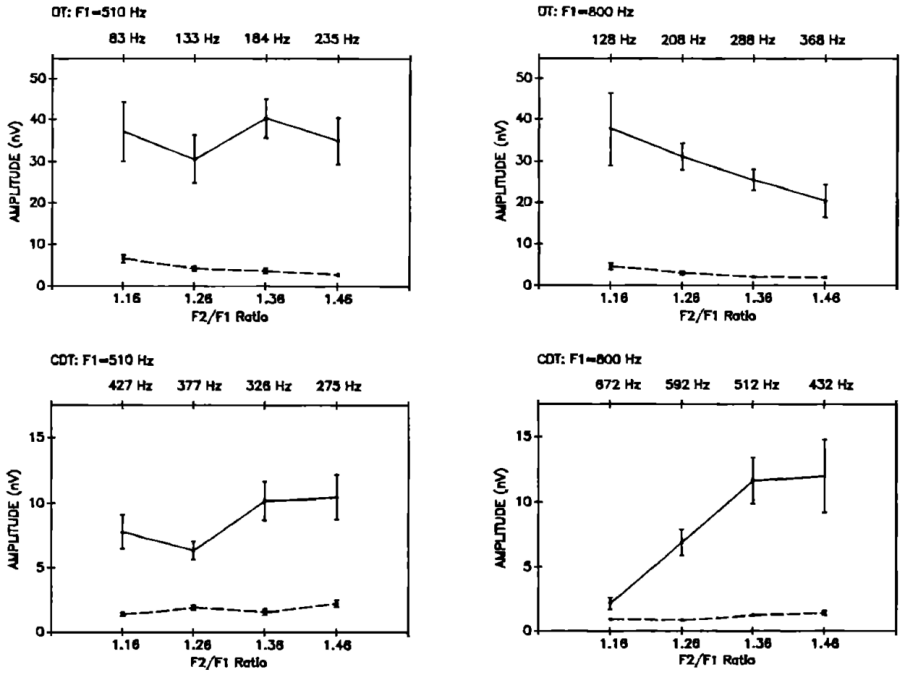


FIG. 4. The solid line shows mean spectral amplitude of DT- and CDT-evoked responses and the associated standard error as a function of f_2/f_1 for ten subjects. Mean spectral amplitude of the no-signal trials of five subjects and the associated standard error is shown as a dashed line.

Figure 1.2: Summary plot of results from Rickman et al. (1991). Their "DT" (difference tone) is equivalent to our QDT here.

exclusively human subjects. Just as responses to two polarities can be averaged, or summed, to yield only even-order responses, they can be subtracted to do the opposite. By subtracting responses to the two polarity conditions one from the other, one obtains only odd-order responses, which include responses to stimulus primaries, along with all higher-order odd-order NFCs, notably the CDT. This study utilized both the addition and subtraction responses in order to compare the CDT and QDT across two f_1 frequencies and four f_2/f_1 ratios for each f_1. Their relevant summary plot is provided in Figure 1.2,

with "DT" (difference tone) being the QDT.

For higher f_1, they note a trend of the QDT amplitude to decrease for increasing f_2/f_1 ratio, but this was not significant. However their CDT for higher f_1 showed significantly increasing amplitude for increasing f_2/f_1 ratio. They also mention that their CDT data is not in accord with two other ways of measuring the amplitude of NFCs from previous work, namely single-unit physiology and psychophysical measures. They point out that the response at the CDT in single fibers in the auditory nerve, measured via spike rate of the fiber whose characteristic frequency is equal to the CDT frequency, decreases as f_2/f_1 increases (Buunen and Rhode, 1978), contrary to their own results. Additionally, psychophysical measures of the CDT decrease as $f_2/f_1 > 1.1$ (J. L. Goldstein, 1970; J. L. Goldstein and Kiang, 1968; Zwicker, 1979). The main psychophysical method involves the subject reporting when they hear the NFC, and when they stop hearing it. The experimenter delivers the two stimulus primaries, as well as a third stimulus component at the frequency of the NFC of interest, but with opposite phase to that of the physiologically generated NFC. When this third stimulus component reaches the correct amplitude and phase, the subject reports that they no longer hear the NFC because it has been fully canceled out.

Another important aspect about the NFCs Rickman et al. (1991) reported is that in general the QDT is of significantly higher amplitude than the CDT in far-field potentials.

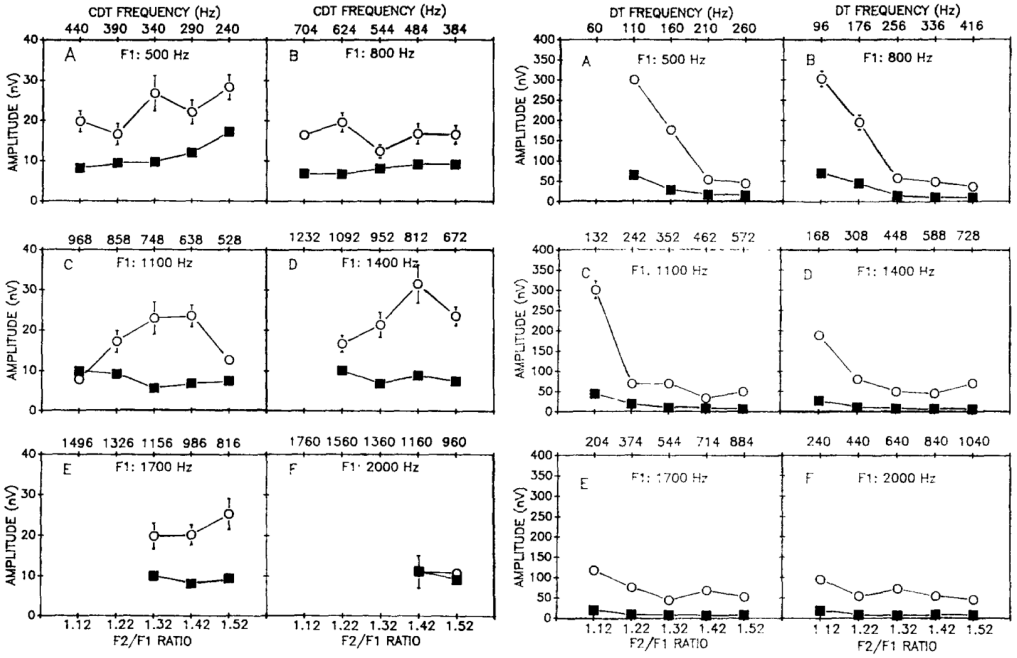


FIGURE 4. Mean amplitude (nanoVolts) of the AEP-CDT as a function of f2/f1 ratio. The error bars are ± one standard error of the mean. The filled squares show the amplitude value for which 95% of the amplitudes from the no-stimulus distribution fall below.

FIGURE 5. Mean amplitude (nanoVolts) of the AEP-DT as a function of f2/f1 ratio. The error bars are ± one standard error of the mean. The filled squares depict the 95% amplitude criteria.

Figure 1.3: Summary plot of results from Chertoff, Hecox, and R. Goldstein (1992). Their "DT" (difference tone) is equivalent to our QDT here.

This would come to be a common theme in the subsequent FFR literature. Chertoff, Hecox, and R. Goldstein (1992) followed up with another study in guinea pigs and expanded on the conclusions from Rickman et al. (1991). This time they found that the QDT indeed significantly decreased in amplitude with increasing f_2/f_1 ratio, and the CDT again significantly increased with increasing f_2/f_1, for lower f_1s. In general the results were more variable at higher f_1s, presumably because it becomes harder for neurons to phase

lock their action potentials and post-synaptic potentials to a given frequency as it becomes higher. This study also confirmed that in general QDTs are larger in amplitude than CDTs in far-field responses. In a single, idealized nonlinear system, nonlinear components of increasing order are increasingly weaker. This difference in amplitudes could be attributed to the fact that the QDT is a second-order NFC while the CDT is third-order, making it inherently weaker than the QDT. However we know the situation is more complicated because this amplitude relationship is not the same for other methods of measurement such as DPOAEs, as Rickman et al. (1991) pointed out. These results are the first to characterize some prominent differences between the QDT and CDT in far-field potentials, strongly suggesting differing mechanisms of generation.

The main results of Chertoff, Hecox, and R. Goldstein (1992) are shown in Figure 1.3. One additional difference between the QDT and CDT can be clearly seen here: The amplitude of the QDT seems to be a pure function of its own frequency. As stated above, the QDT amplitude decreases for increasing f_2/f_1, but $f_2 - f_1$, the frequency of the QDT, also necessarily grows as the ratio grows. We can see that regardless of both f_2/f_1 and the value of f_1 itself, the amplitude of the QDT is only a function of its own frequency. It is roughly at its peak in the range of $90 - 130$ Hz, it drops to about half of that amplitude around $160 - 180$ Hz, and falls steeply from there, being fairly flat at higher than 240 Hz. The CDT on the other hand is clearly affected by both f_2/f_1 and f_1, as its amplitudes

differ for a single value of its own frequency as a function of both f_2/f_1 and f_1.

Research in further characterizing these NFCs has remained slow but steady since the initial comprehensive efforts discussed above. The CDT has been observed resulting from slightly more naturalistic stimuli, such as two-tone approximations of English vowels in, for example, Krishnan (1999) and Elsisy and Krishnan (2008). The QDT was not observed in these studies, nor was it sought after. In fact it would have been impossible to observe this NFC in these two studies because stimuli were delivered in alternating polarity, and the polarity conditions were subtracted from each other, leaving only odd-order responses. But also, in many of the stimuli in Krishnan (1999), the CDT and QDT would be almost the same frequency. If the stimulus duration or the discrete Fourier transform length were not long enough, it would be impossible to tell the two apart.

Although Elsisy and Krishnan (2008) had some limitations, the study is useful for two reasons. One reason is that the stimulus level was systematically varied to map a coarse input/output function of the amplitude of the CDT. The other reason is that the FFR CDT was measured concurrent with the DPOAE CDT to document the differences, if any. The authors found that the FFR CDT grows in a compressive, nonlinear manner as a function of stimulus intensity, whereas the DPOAE CDT only grows compressively for moderate stimulus levels, and grows linearly for both the lowest and highest levels. They explain this by positing that the FFR CDT results partially from cochlear nonlinearity but

also from neural saturation, and also that the FFR CDT originates from only a single place on the cochlear spiral corresponding to the frequency $2f_1 - f_2$. The DPOAE CDT, on the other hand, which will be discussed in detail in its own section, likely results from the summation of amplitude and phase of multiple locations on the spiral: Both the tonotopic place of the CDT and the tonotopic place of f_2.

Another more involved study from around the same time that also compared DPOAEs and FFR NFCs is Bhagat and Champlin (2004). The DPOAE results will be discussed in their respective section, but there were interesting findings with regard to FFR NFCs in this study as well. The overarching question in this study involved the generating mechanisms of both the CDT and QDT, and the question was asked in multiple ways. Most importantly, they looked at the amplitudes of both NFCs as a function of stimulus duration, all other parameters being equal. They found that for the shortest durations, 26 and 51 milliseconds, the amplitudes were not significantly different, although the mean QDT amplitude was higher. As duration increases further, the QDT maintains an amplitude in dB roughly double that of the CDT, and is significant for durations longer than 51 milliseconds. The stimulus frequencies in this case were 500 and 690 Hz. The authors note that a possible explanation of the greater effect of duration on the QDT could indicate multiple neural sources of this NFC, or in general a greater number of sources than the CDT.

Perhaps the primary aspect of the question of generating mechanisms is whether the nonlinearities of the cochlea are solely responsible for NFCs, regardless of the location from which they are measured. Another experiment in Bhagat and Champlin (2004) addressed this question by using dichotic stimuli. In the simple case here in which a stimulus consists of only two frequencies, dichotic means that one stimulus frequency is delivered to one ear, and the other to the other ear. Measures are taken to ensure that there is no acoustic overlap between the inputs to the two ears. The argument is that, if NFCs are generated neurally, they should be documented in the dichotic condition because even though they are not able to interact in either cochlea, they will interact in neurons at higher levels and lead to NFCs. On the other hand, if the cochlea is solely responsible for NFCs, they will not be observed in the dichotic condition.

There is some evidence from invasive animal studies that the QDT is in fact detectable for dichotic stimuli, although it is weak and not detected in every subject (Arnold and Burkard, 1998; Arnold and Burkard, 2000). Bhagat and Champlin (2004) found a dichotic QDT in 3 of 18 subjects, but they did not find a dichotic CDT in any subject. The authors offer several explanations, the most prominent and likely being that the volume-conducted, far-field QDT signal is simply too weak in most mammalian brains to be compared to a direct, invasive recording from individual units or multi-unit responses. The fact that it was detected at all, where the CDT was not detected, may still

offer insights. It shows that it is probably reasonable to say that the cochlea is not entirely responsible for generating even-order nonlinearities.

It should also be pointed out that, from a physiological standpoint, dichotic stimuli are not the best-controlled way of asking the question at hand. This is because, while the method does bypass frequency interaction in the cochleae themselves, it also bypasses the cochlear nuclei on each side. These are the first group of synapses after the hair cells synapse on the auditory nerve, and are located on each side lateral to the olivary complex, where the anatomy from the two sides starts to come together. While there is a very small amount of contralateral efferent connection to the cochlear nuclei (Schofield and Cant, 1996), they for the most part receive only ipsilateral input. This means that in the dichotic condition, they are each only receiving input at the ipsilateral frequency and the frequencies do not have an opportunity to interact at this site of potential nonlinearity. In fact recent research shows that the cochlear nuclei are indeed sites of important nonlinear responses (Laudanski et al., 2010); thus dichotic stimuli may not be the most optimal way of asking whether the cochlea itself is responsible for generating NFCs.

An emerging theme is that the QDT and CDT have different behaviors in many respects. In general for FFRs, the QDT is much stronger than the CDT, often by an order of magnitude or more. Another pattern is that the amplitude of the QDT has an inversely proportional relationship with f_2/f_1 ratio, where the CDT's amplitude has a directly

proportional relationship. The dichotic tests done in both humans and non-humans give weak but real evidence for generation of the QDT and other even-order nonlinearities in the midbrain or rostral to the midbrain, since this is the first location to receive mixed input from both sides. There is no evidence that the CDT is generated neurally, which is consistent with the hypothesis that along with other odd-order nonlinearlities, it is generated by mechanical compression due to the outer hair cells in the cochlea. Odd-order nonlinearities then remain in the response signal that will later be recorded as the FFR as it passes up the auditory brainstem. Neural contribution to even-order nonlinearities from the cochlear nuclei is currently unknown because dichotic stimuli do not test for it; therefore this is an open question that could be addressed invasively or non-invasively. For instance, it is generally accepted that the FFR is a summation of several different sources, but the weighting of those sources varies greatly depending on the electrode montage. In general a lateral montage of one type or another should be more reflective of earlier rather than later brainstem activity (Skoe and Kraus, 2010).

1.3 Laser interferometric velocimetry

A more recent technique for measuring aspects of the auditory system is laser interferometric velocimetry (LIV). Methods of measuring the basilar membrane's motion had been both complex and inaccurate up until the application of this technique in auditory neuroscience. While it is a complicated setup, LIV is the first method to linearly record basilar membrane vibration; in other words, the recorded time series is a "linear function of the target velocity" (Ruggero and Rich, 1991), so that there is no inherent distortion in the recording process itself. This had not been the case with the most popular method previous to LIV, the Mössbauer technique. While also measuring velocity, the Mössbauer technique did so by detecting gamma radiation from a decaying radioactive isotope of cobalt placed on the membrane, which introduces distortion.

LIV measures velocity by instantaneously measuring the Doppler shift of laser light. This technique is invasive and requires sedation and is thus not performed on humans. A common animal used in these studies is the chinchilla. A very small, extremely reflective glass bead is placed on the basilar membrane after surgery. A laser is then pointed directly at the bead, reflecting the light back into the same device emitting the beam. If the membrane is in motion, such as when it is resonating to sound, movement of the bead will cause a Doppler shift in the frequency of the laser light as it is reflected back into the device, similar to the Doppler shift of the frequencies of a train whistle as it approaches,

then recedes from, a listener. The apparatus records the instantaneous difference between the emitted light frequency and the Doppler-shifted light, which is the measure of instantaneous velocity. We note that because velocity is the derivative of displacement, we are not strictly measuring "motion" here, but rather its derivative.

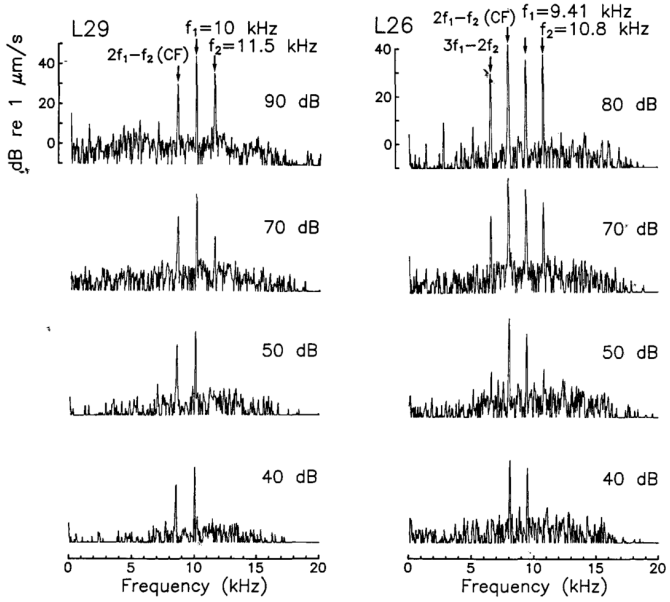


FIGURE 1 Frequency spectra of responses of basilar membranes L29 (left) and L26 (right) to two-tone stimuli, measured using laser velocimetry at the basal turn of the cochlea. Spectra were obtained by Fourier transformation of responses to equal-level tones (40-90 dB SPL) with frequencies chosen such that $2f_1$-f_2 corresponded to the characteristic frequency (8.5 kHz in L29 and 8 kHz in L26) of the basilar membrane location under measurement. The spectra show, in addition to peaks at the primary frequencies (f_1 and f_2), a prominent distortion peak at $2f_1$-f_2 in both animals, as well as at $3f_1$-$2f_2$ in L26. Note the disappearance of the response at frequency f_2 at the lowest stimulus levels.

Figure 1.4: Some frequency-domain data from Robles et al. (1990) taken from two animals. Stimuli were two-tone complexes. These responses are recorded from the place on the membrane whose CF is equal to $2f_1 - f_2$.

This method was first described, as applied to detecting NFCs, by two labs at a

conference in Madison in 1990 (Nuttall, Dolan, and Avinash, 1990; Robles et al., 1990), and was first detailed in the literature in Ruggero and Rich (1991). Some typical LIV results from two-tone stimuli (two sinusoids) are shown in Figure 1.4. Ruggero and Rich (1991) state that the noise floor of individual responses was much too high: roughly the same level as responses to the primaries, except at the highest levels of stimulation. Thus thousands of responses to identical stimuli are averaged, as is the case with recording FFRs, to produce a plot such as Figure 1.4.

It is common to look at the CDT, at $2f_1 - f_2$, when using LIV, and this is partially what the two papers at the Madison conference did. Often the stimuli are designed so that $2f_1 - f_2$ is the CF of the place on the membrane the glass bead was set. This was the case for Figure 1.4, which shows a very prominent peak for the CDT in both animals. We also see another odd-order NFC below the CDT for high stimulus levels in one animal. We will see that in subsequent research, many additional odd-order NFCs such as this become the rule rather than the exception. In this case we see a quintic distortion product, at $3f_1 - 2f_2$.

Robles et al. (1990) and Nuttall, Dolan, and Avinash (1990) both point out that growth in amplitude of the CDT NFC as a function of stimulus level is compressive, whereas growth in response to the stimulus primaries was roughly linear. In these experiments, the location of measurement is the place on the membrane where CF equals the CDT frequency. So we can generalize by saying that growth in the response amplitude

of a given frequency around the CF of the location of measurement is nonlinear, and growth in the amplitude of frequencies significantly off from CF is linear. For small f_2/f_1 ratios when the primaries are very close to the CDT, however, responses to the primaries will be compressive as well, especially at high stimulus levels, presumably falling into the same auditory filter as the CDT. Both studies also report that the CDT becomes weaker as the f_2/f_1 ratio increases, though the growth functions for different ratios have roughly the same slope.

As researchers improved in their usage of LIV, data became cleaner and clearer and SNRs got better. Figure 1.5 shows that the basilar membrane appears to be behaving like a generalized odd-order nonlinear system, with many higher-order NFCs appearing relatively equally above and below the responses to the primaries. In the case of a small f_2/f_1 ratio, responses to the primaries are always of greater amplitude than the NFCs, but for a larger f_2/f_1 ratio responses to the CDT are always greater than those to the primaries (Figure 1.5B). A possible interpretation of these differences is that, in the case of $f_2/f_1 = 1.05$, the primaries fall into the same auditory filter and thus have more opportunity to interact nonlinearly, whereas they do not fall into the same filter in the case of $f_2/f_1 = 1.25$. In other words, for $f_2/f_1 = 1.05$, the primaries are unresolved by the auditory system, but for $f_2/f_1 = 1.25$, the primaries are resolved.

Robles et al. (1997) also studied the effect of varying the amplitude of only one of the

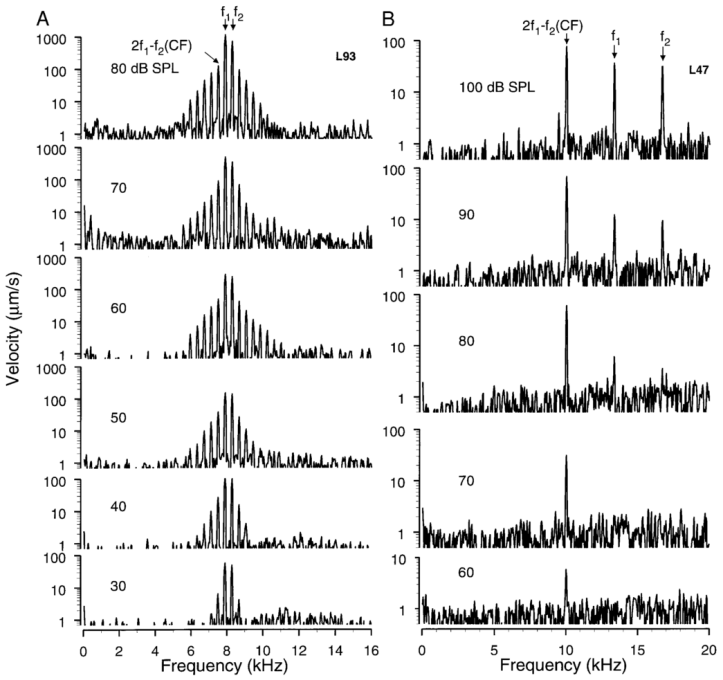


FIG. 2. Effect of frequency spacing on the spectrum of BM responses to tone pairs (f_1, f_2, chosen so that $2f_1 - f_2$ = CF). *A*: spectra of BM responses to pairs of tones with closely spaced frequencies near CF ($f_2/f_1 = 1.05$). The equal-level primary tones had $f_1 = 7.89$ and $f_2 = 8.28$ kHz, such that $2f_1 - f_2 = $ CF (7.5 kHz), and were presented at 30–80 dB SPL. *B*: spectra of BM responses to pairs of tones with widely spaced frequencies, such that $f_2/f_1 = 1.25$ and $2f_1 - f_2 = $ CF (10 kHz), presented at 60–100 dB SPL.

Figure 1.5: Some frequency-domain data from Robles et al. (1997) taken from one animal. Stimuli were two-tone complexes. These responses are recorded from the place on the membrane whose CF is equal to $2f_1 - f_2$. Panel A shows responses for $f_2/f_1 = 1.05$ and panel B shows responses for $f_2/f_1 = 1.25$.

two primaries on the amplitude of the CDT at $2f_1 - f_2$. For a variety of fixed f_2 levels, they measured the level of the CDT as a function of a varying f_1 level. Consistently, the largest level of nonlinearity is found when the level of f_1, closer in frequency to the CDT than f_2, is 5 or 10 dB higher than f_2. The opposite test was also done, in which the level of f_2 changes for a fixed f_1 level, and the same pattern was found, in which an f_2 level 5 or 10

dB below the fixed f_1 level made for the largest amount of nonlinearity. They also point out that, when expressed as a ratio, the greatest amount of this odd-order nonlinearity is found at the lowest stimulus levels. Just as responses to the primaries grow compressively for greater stimulus levels, the responses at the CDT and other odd-order NFCs grow compressively with respect to the responses to the primaries.

What is conspicuously missing from the LIV discussion so far is any mention of even-order NFCs, mostly the QDT. In fact it appears that these components are indeed absent from the basilar membrane responses that have been reported in the literature. In these experiments, the glass bead is placed on the membrane at a location that is most convenient for the experimenters; the CF of that location is then determined and stimuli are designed around that CF, most commonly, as has been mentioned, two-tone stimuli whose CDT is CF. The CFs we have seen thus far have been between roughly 8,000 and 17,000 Hertz. The cochlea is a spiral structure so naturally the more convenient locations will be less apical and more basal, and since the frequency gradient on the membrane goes from low at the apex to high at the basal end, it makes sense that the CFs used in these experiments would be relatively high. The data certainly show that the QDT and other even-order NFCs are not being generated on the membrane at or near the CDT frequency location (which is often close to the location of the primary frequencies of course, in the case of small f_2/f_1 ratios), and if they are generated elsewhere on the membrane, the data

show that they are not being propagated to this location. This does not however prove that they are not being generated on the membrane at all. One would need to position the bead at much more apical locations to be near typical QDT frequencies, and this is very inconvenient or impossible given the setup of LIV.

Nuttall and Dolan (1993) asked the question directly whether the QDT is present in the cochlea. They measured the cochlear microphonic (CM) at the round window, hair cell potentials with a glass microelectrode, as well as basilar membrane velocity with LIV, all simultaneously. Using two-tone stimuli, they found that for a variety of $f_2 - f_1$ frequencies, absolute primary frequencies, and levels, the QDT was present in the CM and in the potentials of the hair cells, however they did not observe it in basilar membrane velocity at all. A confounding factor, as explained above, could be that the glass bead was placed at roughly the location where CF is close to the primary frequencies as opposed to the $f_2 - f_1$ frequency, however this was also the location at which they measured hair cell potentials and found the QDT prominently. They measured both inner and outer hair cells, and interestingly most of the QDT power was coming from the inner, rather than the outer hair cells. The inner hair cells are commonly taken to be more or less linear transducers, with the outer hair cells otherwise being known as the main culprits of nonlinearity in the cochlea, including its extremely sharp tuning, compressive nature, and exquisite sensitivity.

Another interesting aspect of this particular instance of the QDT is its amplitude as a

function of stimulus level. It is nonmonotonic, peaking at a stimulus level of only 45 dB SPL, and falling again at all higher levels. This is quite uncharacteristic of what we have seen in, for instance, the FFR literature that has reported on NFCs. The authors note however that the absolute potentials seen for the QDT are quite high in the first place, being in some cases 50 times the power at the primary frequencies. Nevertheless the nonmonotonic growth function is quite at odds with the robust QDT consistently seen in both the FFR and in DPOAEs, discussed below. The authors acknowledge this and are not confident that the QDT power seen in the IHCs is mostly responsible for the physical QDT seen in the ear canal or the electrical QDT observed in scalp-recorded responses. The QDT DPOAE in particular has a much different and more interesting set of behaviors than the CDT DPOAE, and the authors suggest that the former may be a result of efferent neural responses as opposed to intrinsic cochlear dynamics, a possibility also raised in the modeling efforts of Lerud et al. (2014).

Little other published LIV work either seeks out or shows evidence for even-order nonlinearity in basilar membrane velocity. The fairly comprehensive Robles et al. (1997) report mentions that they witnessed occasional, unreliable responses that seemed to be at $f_2 - f_1$ or its harmonics, but they attribute this data to equipment-induced distortion, and only report on the odd-order NFCs in that paper. Only single sinusoids were used as stimuli in Cooper (1998) and responses were found at $2f_1$ and other harmonics, so this may

be attributed to even-order nonlinearity, but seems to be a different mechanism than the more reliable type of two-tone nonlinearity such as that observed in Figure 1.5.

One consistent finding from LIV that contrasts it from the FFR is that the CDT amplitude is inversely proportional to f_2/f_1 ratio. This is perhaps more intuitive, since one would expect the compressive nonlinearity to weaken as the two stimulus frequencies fall into different cochlear filters. But the main feature of responses obtained through LIV is the lack of even-order nonlinearities. While the velocity of the basilar membrane does not appear to contain the QDT, it is still possible to record the QDT from the cochlea through electrical methods, namely the CM and hair cell potentials. Recorded in this way, the QDT is prominent, and the possibility exists that the efferent olivocochlear system may be responsible for it, since both the medial and lateral portions of that system synapse directly on the hair cells. Hence if the cochlear nucleus or the olive itself were responsible for the generation of the QDT, it would make sense to be able to detect it electrically within the cochlea for this reason.

1.4 Otoacoustic emissions

Because of the nonlinearity of the cochlea, particularly that of the outer hair cells, the ear actually emits sound under a variety of circumstances. These pressure waves, which originate in the cochlea and which are detectable with a sensitive microphone in the ear canal, or even occasionally with another human ear, are called otoacoustic emissions (OAEs) and come broadly in four types. Stimulus-frequency OAEs are emissions which replicate a stimulus well, usually a single sinusoid. Transient-evoked OAEs are broadband OAEs triggered with a quick broadband stimulus like a click. Spontaneous OAEs vary widely between individuals, but when they exist they are present without any immediate auditory stimulus. And distortion-product OAEs, the type that interests us here, replicate a multi-component stimulus but also include various NFCs.

OAEs were first reported by Kemp (1978), with Kemp (1979) already reporting DPOAEs specifically. The latter study used several different tone pairs with a variety of ratios, detecting the CDT every time through a simple frequency analysis of the pressure recorded in the ear canals of human subjects. While this study only looked at the CDT, Kim et al. (1980) also found the QDT DPOAE. Even though the DPOAE is a non-invasive measure, the researchers in this study were working with cats because they were also gathering auditory nerve spiking data. While data from humans and cats are not entirely comparable, this study reports that the QDT is roughly an order of magnitude smaller

than the CDT, in multiple animals and for multiple tone pairs.

Several prominent differences between the QDT and CDT are apparent in the case of DPOAEs, which may shed some additional light on the possible generating mechanisms of different types of NFCs. One area that has been explored is the effect of contralateral stimulation on ipsilateral DPOAEs. Presumably to the extent that this occurs, it is an effect mediated by the medial olivocochlear system (MOC). After the eighth cranial nerve synapses on the cochlear nucleus, fibers decussate and then synapse again on the superior olive. From here, there is a small bundle of efferent fibers that project back to the ipsilateral side, synapsing directly on the outer hair cells.

One general result that has been replicated a number of times is that contralateral wideband noise at sufficiently high levels has a large effect on the ipsilateral QDT DPOAE, but not its CDT counterpart. Wittekindt et al. (2009) report for instance that, for a two-tone complex at 71 dB SPL, there is significant suppression of the QDT with contralateral noise as low as 40 dB SPL. The suppression further increases with higher noise levels. The phase of the QDT also leads, with respect to its phase with no contralateral noise, with increasing noise level. The CDT by contrast shows none of these features in this study.

One of the first papers to look at these issues showed the same trends, but with more manipulations and slightly more mixed results (Kujawa, Fallon, et al., 1995). They showed

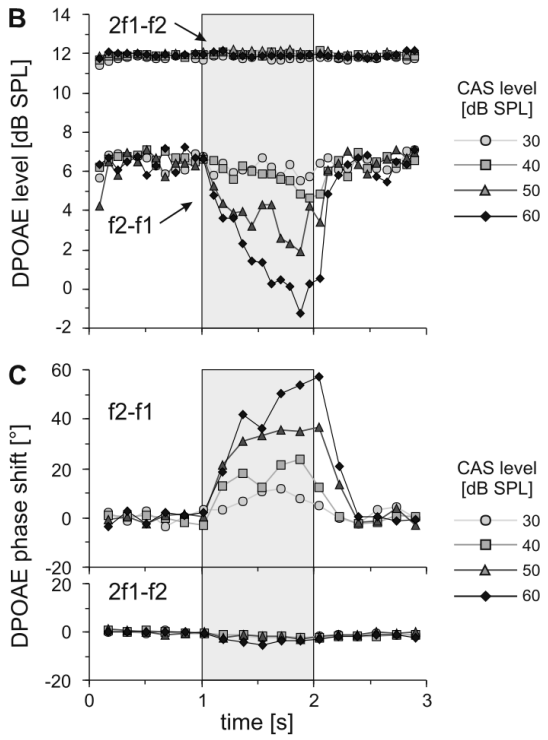


Figure 1.6: Human DPOAE data from Wittekindt et al. (2009). A clear effect of contralateral acoustic stimulation (CAS) on the QDT is visible, and a lack of effect on the CDT is also apparent.

the suppression of ipsilateral QDT DPOAE with contralateral wideband noise, but also showed suppression of the CDT, although it was smaller than that of the QDT. Direct comparisons between the two studies can't be readily made because of their differences, not the least of which is that Kujawa, Fallon, et al. (1995) used guinea pigs so that they could conduct invasive measurements and manipulations, where Wittekindt et al. (2009) used human subjects. One of the invasive manipulations in the former study was the application

of tetrodotoxin (TTX) to the MOC system. This had a clear effect of eliminating the contralateral suppression; specifically, both the QDT and CDT remained at their original levels with and without contralateral noise. Their final manipulation was to section the MOC completely. This had two effects, interesting in their difference from the effect of TTX. Sectioning the MOC at the midline eliminated the contralateral suppression. The other effect was that the level of the QDT, now the same with contralateral stimulation as without it, shifted down by about 30%. This was not however the case for the CDT; it maintained its pre-section level.

Some representative human data from Wittekindt et al. (2009) is shown in Figure 1.6. A clear effect of contralateral noise at higher levels is evident on both the amplitude and phase of the QDT, where there is no effect on the CDT. Althen et al. (2012) took similar measurements, this time from the gerbil. The result that contralateral noise affects the QDT but not the CDT was replicated, but the direction of the effect on the QDT was not consistent even within individuals. The tendency was an enhancement rather than a suppression as seen in other studies. The authors suspect that the difference is due to the species, and attribute it to different operating points on the cochlear transfer function.

The main difference between the behavior of these NFCs in DPOAEs compared to FFRs is that the QDT is smaller than the CDT in general. This seems expected for the same reason it is expected from the basilar membrane directly, which is that cochlear

nonlinearity is assumed to be largely compressive, generating chiefly odd-order NFCs. Therefore the question as to why there is a mechanical response at the QDT at all is interesting. If even-order nonlinearities are generated by the olive or beforehand and then create the QDT DPOAE through the olivocochlear system, it would seem that the QDT should have fully disappeared when Kujawa, Fallon, et al. (1995) sectioned the MOC. There was a significant drop in its amplitude, but it did not disappear. However, Kujawa, Fallon, et al. (1995) only sectioned the MOC, with a cut at the midline. This leaves the lateral olivocochlear system (LOC) intact, as this portion of fibers does not decussate. In the LOC system, the ipsilateral olive receives input from the cochlear nucleus, and projects fibers directly back to the ipsilateral cochlea, specifically onto the inner hair cells. If the LOC were responsible for a significant portion of the mechanical QDT, the hair cell recordings reported in Nuttall and Dolan (1993) may make some sense. They reported most of the electrical QDT in the potentials from the inner hair cells, with little response from the outer hair cells. Since this is where the LOC system synapses, those potentials seem to be located where they should be.

1.5 Concluding remarks

The cochlea is usually understood to have nonlinear properties that are important to auditory perception, but given the differences between the three types of measurement explored here, there can be little doubt that higher auditory structures contribute to NFCs as well. We have seen, for instance, that the FFR contains many even-order NFCs, most prominently the QDT, whereas no even-order nonlinearity seems to be physically present on the basilar membrane at all. Contralateral stimulation can significantly affect the QDT DPOAE while not affecting odd-order DPOAEs, indicating that even-order distortion may be partially generated neurally and propagated back to the cochlea through the efferent system.

The prominent NFCs do not seem to be direct correlates of perceived pitch (Gockel et al., 2011), however their presence at certain levels is often indicative of healthy auditory processing. Smalt et al. (2012) measured the scalp-recorded FFR and paired it with a behavioral task that measured fundamental frequency difference limens for complex stimuli in various levels of noise. The tones were missing fundamental complexes made up of several high harmonics, and the noise was lowpassed to have a bandwidth from 0 Hz to right below the primary frequencies. The authors showed that, for the higher levels of noise, the odd-order NFCs, including the CDT, right below the responses to the primaries were degraded, and that this degradation predicted the degradation in the fundamental

difference limen. Interestingly, while the CDT was negatively affected by the noise, the QDT remained unaffected. The behavioral component to this particular study is potentially informative as to the practical purposes of studying the physiology of NFCs.

As we have seen, certain NFCs particularly in FFRs and DPOAEs are extremely reliable, such that they are used clinically to assess the health of the auditory system. While this in itself is a good indicator that NFCs are an aspect of physiology that is important for perception, results in recent years particularly for the FFR have provided more precision. While the NFCs in the FFR do not directly represent perceived pitches, they may nevertheless indicate them. Wile and Balaban (2007) measured the residue pitch of shifted frequency complexes along with the FFR. They showed that the pitch was predicted by an average of the QDT and CDT if the average is weighted by the respective amplitudes of the two components. They also used bandpassed noise in the region around these NFCs, which as we have seen may likely have affected both the perception and the CDT. But if this result is more or less true and can be replicated, it is a good window into the relationship between the essential nonlinearity in auditory physiology and the larger realms of perception and behavior.

Chapter 2

Residue pitch perception of shifted frequency complexes

2.1 Introduction

Pitch is a perceptual phenomenon rather than a physical one, and much effort, particularly over the last century, has been devoted to finding and characterizing the various physical causes of the percept. Pitch is described by the American National Standards Institute as "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (de Cheveigné, 2005). As a psychological aspect of auditory perception, it is analogous to the concept of pulse, or tactus, on the timescale of musical

rhythm: It may be well-predicted or inferred from the physical sound itself, but this is not always the case.

Important to the discussion below, pitch may be thought of in both an absolute and relative sense. For instance, one can say with confidence about many sounds that they have a more or less single, unambiguous pitch which can be defined as a frequency, usually in units of Hertz (Hz), or cycles per second. This definition would take place in a psychological experiment in which a listener measures the pitch frequency of the sound for the experimenter, usually by matching the pitch of an objective sound, often called the comparison, to the pitch of the experimental sound, often called the standard. The pitch of the standard has then been psychologically defined as the frequency of the comparison that the subject ended up on. For the purposes of the experiment, the comparison is taken to have an objective pitch; this is sometimes appropriate, but ultimately since pitch is inherently psychological, there cannot be any sound with a truly objective pitch. Indeed a pure tone, a single sinusoid, would seem the most rational choice for a comparison tone and has frequently been the choice, surely having the pitch of its frequency. However it has been shown in more recent times that a sinusoid does not have the same pitch as a stack of its own harmonics (G. A. Moore and B. C. J. Moore, 2003; Walliser, 1969).

Pitch can also be conceived relatively. For instance, one can imagine wideband, noisy sounds for which it may be impractical to attempt to assign a specific pitch frequency.

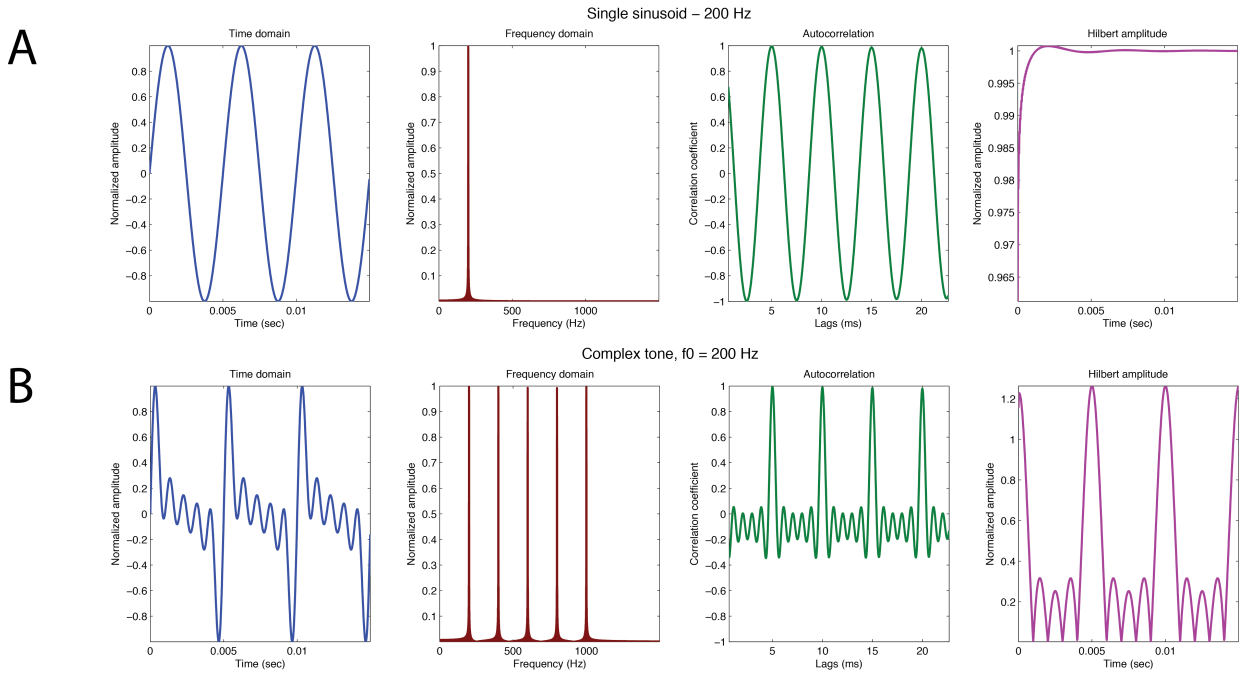


Figure 2.1: Summary plot demonstrating basic characteristics of a pure tone, or a single sinusoid (A), and a complex harmonic tone (B).

However, it may nevertheless be the case that given two sounds of this type, listeners would reliably rate the pitch of one higher than the other. Even without a label of a frequency, pitch is still a psychological reality for many types of sounds. Also important for the discussion below is that some sounds may have a pitch perceptual correlate that is bistable or otherwise multistable, and that may undergo perceptual hysteresis; that is, a pitch correlate of the sound may have a definite frequency, but there may be more than one that can be reliably measured from listeners, and which is selected by the listeners may depend

on such context effects as recency of certain other sounds or other concurrent sounds.

Many researchers attempt to develop comprehensive models to predict the pitch frequency/frequencies of arbitrary sounds, either by way of modeling the relevant auditory physiology or simply processing the sound in some way. Two conclusions that have become increasingly accepted as a result of these efforts are: The sound need not contain the pitch frequency in its spectrum, and the relevant physiology need not contain the pitch frequency in its spectrum. The latter conclusion, despite many experiments and models, has only become generally accepted within the past decade. The following is a brief empirical review of the facts surrounding a phenomenon called pitch shift of the residue.

2.2 History and missing fundamental

Missing fundamental perception has been understood since the 18th century. When only higher harmonics (integer multiples) of a fundamental frequency are present in a stimulus, but not the fundamental itself, a pitch is nevetheless typically perceived at the missing fundamental frequency. Figure 2.1 compares some aspects of a single sinusoid (panel A) and what is usually called a "complex" tone, consisting of multiple harmonics of the original sinusoid (panel B). Figure 2.2A then demonstrates a missing fundamental

stimulus, as only harmonics 3, 4, and 5 of the missing fundamental 200 Hz are present. The stimulus in Figure 2.2A has the same pitch as that in Figure 2.1B.

Giuseppe Tartini first pointed out in 1754 that a person's auditory system must somehow be adding combination tones to their perception in the context of music (Hudspeth et al., 2010), in particular the organ. He noted that an organ need not produce, for instance, a desired very low note for that note to be perceived; it need only produce multiple higher harmonics of the desired note, and the proper note will typically be perceived. This had practical implications, since an organ requires increasingly larger pipes to physically produce increasingly lower notes.

The 19th century saw the first attempts to explain this phenomenon, and the beginnings of competing theories of pitch perception in general. Since the purpose of the present work is to review what is known about a perceptual phenomenon, detailed accounts or critiques of theories of pitch perception will only be of concern when directly relevant. Seebeck (1841) laid out the first example of what would generally come to be called a "temporal" theory of pitch perception (Schouten, 1940b), and used it to explain the perception of pitch at the missing fundamental of a complex tone. A temporal theory relies on both the amplitude envelope and fine structure of the waveform to explain the pitch; as can clearly be seen by comparing figures 2.1B and 2.2A, a missing fundamental stimulus has peaks in its amplitude envelope at the same places as the full-spectrum complex, every

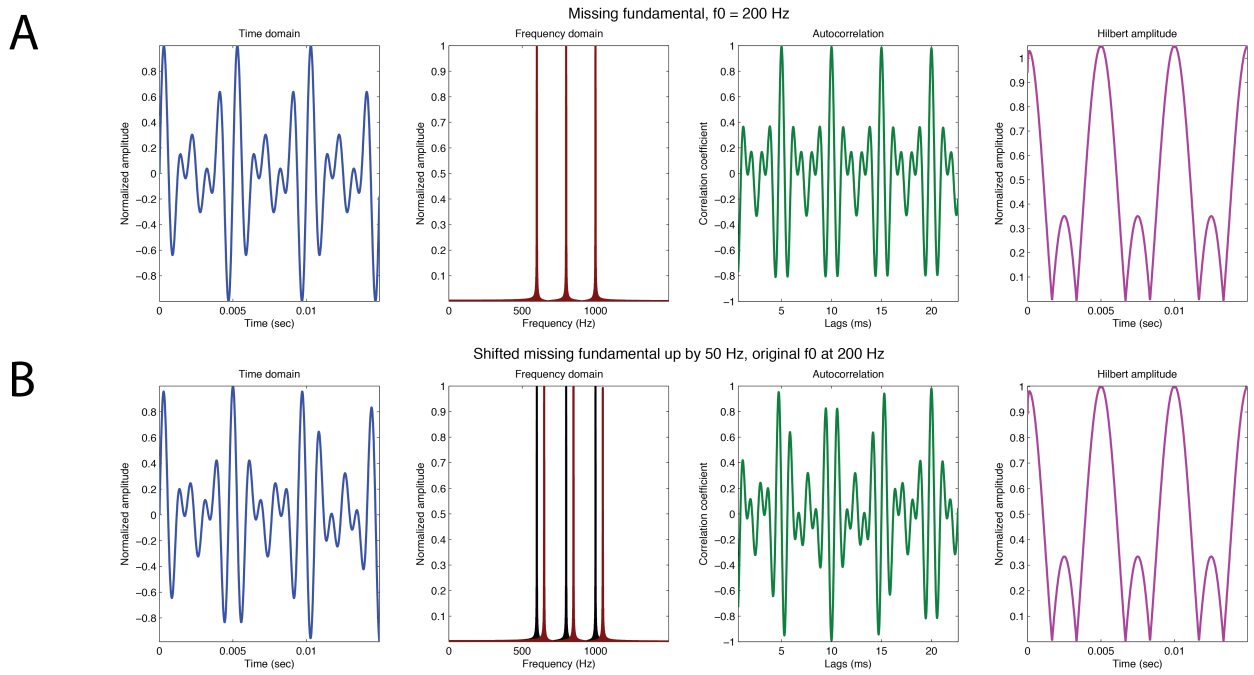


Figure 2.2: Summary plot demonstrating basic characteristics of a harmonic missing fundamental tone (A), and a frequency-shifted missing fundamental tone (B). The original three frequencies (in black) are only shown in B for reference.

5 milliseconds, the reciprocal of which is 200 Hz, the pitch frequency.

The second general type of pitch perception theory is called a "spectral" theory. In many ways, that were known in the 19th century and are still generally accepted today, the inner ear, mostly the cochlea, acts as a frequency decomposer. This is because the basilar membrane has a logarithmically-varying thickness from base to apex, such that each place on the membrane is preferentially sensitive to a specific frequency. In this way each place

on the membrane acts like a resonator, or a bandpass filter. The complete story about cochlear mechanics is of course much more complex and nuanced, but this has always been an attractive starting point for people theorizing about the auditory system, as it appears that the first step in auditory perception is something like a Fourier analysis, giving the spectrum of the input in some form. The first well-developed theory based on this, written as a counterpoint to Seebeck, was Ohm (1843). The spectral presence of the fundamental of a complex tone was taken to be the reason for the pitch at that frequency, so an account needed to be given of how it got there if it was not in the stimulus, as is the case with missing fundamental stimuli. von Helmholtz (1863) expanded on Ohm's work, agreeing with him, and further developed a spectral theory of pitch perception (Schouten, 1940b), in which missing fundamental perception was explained as the result of a nonlinear distortion product at the fundamental frequency caused at some early stage of the auditory system.

It turned out that Helmholtz was right that the auditory system, including the cochlea, is responsible for generating a wide variety of nonlinear frequency responses, including the one that would predict the pitch of such missing fundamental stimuli as Figure 2.2A. The question remained, however, whether this spectral correlate of auditory processing is indeed responsible for the pitch perception. There are various manipulations one can conceive of to test this hypothesis, and perhaps the simplest and one of the most intriguing is exemplified in Figure 2.2B. In a missing fundamental stimulus like that of

Figure 2.2A, in which all harmonics that are present are consecutive, the pitch frequency is at the missing fundamental, 200 Hz, which can be calculated simply by noting that the difference between one harmonic to the next is 200 Hz. This common difference frequency also controls the frequency of the amplitude envelope, also pictured. If one shifts the three present harmonics up by a common amount, 50 Hz in the case of Figure 2.2B, we have an interesting scenario. The frequencies present are no longer harmonics of 200 Hz, and the ratio between one harmonic to the next has been changed. Yet the difference between successive harmonics is still 200 Hz, which means that the amplitude envelope is identical to the non-shifted case, Figure 2.2A. The question here of course is whether this sound has a pitch, and if so, what it is. If it has a pitch, Helmholtz and other spectral theories should place it unchanged, at 200 Hz, since the missing fundamental pitch was predicted on the basis of a nonlinear difference tone.

Schouten (1940a) introduced the nomenclature and concept "residue" pitch, which will be used for the remainder of the present work. Residue pitch simply generalizes the notion of missing fundamental perception; if, for instance, the above described complex has a pitch, whether at 200 Hz or something else, it no longer makes sense to refer to it as a "fundamental" because it is not the true fundamental frequency of the frequencies present. Schouten describes the residue pitch as a "collective perception" of the stimulus and attributes it at least partly to the fact that multiple higher harmonics often fall into the

same auditory filters, whose bandwidths are determined in psychophysical experiments. These are usually referred to as unresolved harmonics. In the decades to come it would be fleshed out that residue pitch is also reliably measured when the frequencies in the stimulus are lower and all fall into their own auditory filters; these are referred to as resolved.

And in fact Schouten (1940a) was the first effort to measure the situation described above, the pitch shift of the residue. It was the last topic discussed in the paper, as a kind of aside, and was oddly printed in a smaller type even though it was three paragraphs. In this description he nevertheless lays out the basic facts that would quickly come to be more precisely understood. A shift of this type results in a residue pitch which is neither the difference tone nor the difference tone shifted up by the same amount as the stimulus frequencies; rather the pitch is in between these two values, and how much it moves is inversely proportional to the harmonic numbers of the frequencies in the stimulus. Thus the pitch shift of Figure 2.2B is some number less than 50 Hz, and if it had been harmonics 9, 10, and 11 instead of 3, 4, and 5, the shift would be even less, though still present. The pitches were determined in probably the only way which would be technologically feasible for the time: The shifted stimuli, the standards, are generated with amplitude modulation telephony equipment, and the comparison stimuli are harmonic complexes whose frequencies are manipulated by finely adjusting the speed of a gramophone record. Shifting the comparison complex in this way has the effect of multiplying each present frequency by

a common amount, keeping the complex harmonic. Thus the comparison is between a harmonic sound and the sound in question, which when shifted is no longer harmonic.

The timbre of sounds like these is also discussed in Schouten (1940a) and he points out that it changes dramatically and becomes more metallic with the shift. Sounds like these are often called "anharmonic" or inharmonic, because they are not composed of consecutive harmonics of any frequency, and he points out that the peculiar timbre makes sense since this is the situation with many metallic objects, such as bells. de Boer (1956), who was Schouten's student, mentions that one person documented the core aspects of this phenomenon, the pitch shift and the timbre alteration, as early as 1929; Fletcher (1929) dedicates a single paragraph to the shifted sounds, noting only that the perception of pitch is not destroyed for relatively small shifts but does in fact move with the shift, and that the timbral result of the shift is that the sound "loses its musical character". In fact Schouten (1940a) also credits Fletcher for being the first to describe the phenomenon in a short footnote, but refers instead to Fletcher (1924) which contains the exact same paragraph as the later book, and was likely the first time the statement appeared in print. According to the byline of the latter paper, Fletcher was working at the laboratories of the American Telephone & Telegraph Company at the time, shortly before that company merged with Western Electric to form Bell Laboratories, both of which institutions Fletcher also worked as a researcher at.

2.3 Characterization of the shifted residue pitch

While Fletcher moved on to a variety of other topics, Schouten continued his work on pitch shift of the residue, culminating in the much more systematic Schouten et al. (1962). Here the authors measured the pitch of a variety of shifted missing fundamental tones. The generation of standard and comparison stimuli was similar to that of Schouten (1940a), but with slightly less cumbersome technology. The comparison stimuli, which, as above, are all harmonic complexes whose fundamental is known and can be controlled by the subject, were pre-generated on magnetic tape and the subject controlled the fundamental by controlling the speed of the tape. Generation of the standards is described in more detail than earlier reports.

All standards (and all comparisons) are three-component complexes. For the main experiment in Schouten et al. (1962), the components of each standard are always separated by 200 Hz. Standard complexes which are harmonic have a middle harmonic number, n, at $n = 7$ through $n = 12$. Thus the frequencies in the complex for $n = 7$ are 1200, 1400, and 1600 Hz, etc. The rest of the standards, which will be shifted, can be described as starting from one of these six harmonic complexes, and shifting the complex up or down in increments of 50 Hz. These stimuli lend themselves nicely to studies of pitch perception, and they were also relatively convenient to generate for the time period. The experimenters used two oscillators, one of which served as an amplitude modulator of the

other. A signal $s(t)$, where

$$s(t) = \Big(1 + m\cos(2\pi gt)\Big)\sin\Big(2\pi ft\Big) \tag{2.1}$$

is an amplitude-modulated carrier sinusoid at f Hz modulated sinusoidally at a frequency of g Hz. Assuming the modulation depth m to be 1 for convenience, Schouten et al. (1962) note that the equation for the signal their oscillators are generating can be rewritten in a way that makes its spectrum more obvious. For a harmonic case with n an integer multiple of g,

$$s(t) = \frac{1}{2}\sin\Big(2\pi(n-1)gt\Big) + \sin\Big(2\pi ngt\Big) + \frac{1}{2}\sin\Big(2\pi(n+1)gt\Big), \tag{2.2}$$

and by adjusting the frequency f Hz of the carrier by Δf, we have the general form for a shifted frequency complex

$$s(t) = \frac{1}{2}\sin\Big(2\pi\Big((n-1)g + \Delta f\Big)t\Big) + \sin\Big(2\pi\Big(ng + \Delta f\Big)t\Big) + \frac{1}{2}\sin\Big(2\pi\Big((n+1)g + \Delta f\Big)t\Big) \tag{2.3}$$

whose spectrum clearly consists of the carrier frequency f Hz as well as two sideband components $f \pm g$ at half the amplitude of the carrier. As stated above, in this study

stimuli were generated for $n = 7$ through $n = 12$, multiple Δf in increments of ± 50 Hz until the pitch perception was lost, and with $g = 200$ Hz. Comparison stimuli, which were all harmonic, matched the n of the standards at a given point, so that the standard and comparison were always in the same spectral range. Results from the study are in Figure 2.3.

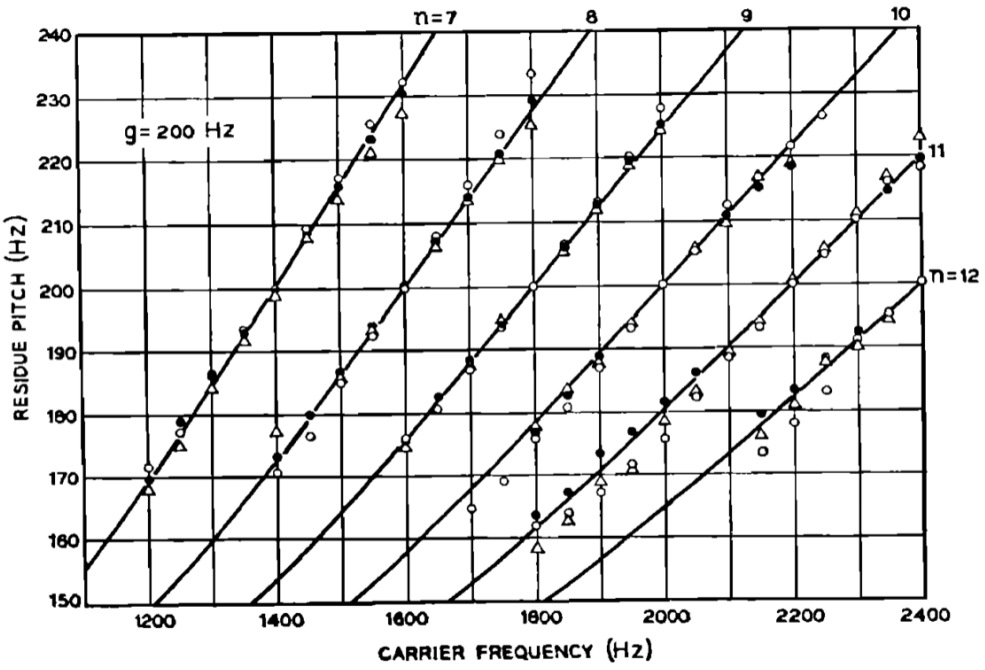


Figure 2.3: Summary plot of pitch shift results of Schouten et al. (1962), reproduced in Schroeder (1966). Each of the three data shapes indicates the mean of 12 trials from one of three subjects. The lines are not best fit lines; rather they represent the pitch shift model described in Schroeder (1966), but which is very close to the lines of best fit.

On the x-axis is the carrier frequency of a given stimulus, which we can remember is

simply the middle frequency of the three-frequency complexes, and on the y-axis are the perceived pitches from three subjects. One pattern we can note right away is the basic tendency that was pointed out in Schouten (1940a), which is that the pitch shift is not as great as the frequency shift, and the shift is less the higher the n. If we have the pitch frequency p and the pitch shift Δp, then this tendency can be approximated by the simple model

$$\Delta p = \frac{\Delta f}{n}. \tag{2.4}$$

An interesting fact shown nicely by Figure 2.3 is that the same physical stimulus is capable of eliciting two, three, or even four different pitches, even within this single study. If one picks a point on the x-axis, this refers to a single stimulus, since the x-axis is f and g is being held constant. Thus when the lines of different n overlap, this means that the same stimulus can have multiple pitches. The experimental design of this study deliberately started, for a given n, with the harmonic situation and then moved steadily up or down, to track the line as far as it could go. This results in the finding that, for instance, the complex made up of 1600, 1800, and 2000 Hz, a harmonic complex, can have a pitch of 200 Hz, the missing fundamental, ≈229 Hz, ≈177 Hz, or ≈161 Hz, depending on the recent perceptual context. Approximations were made by visually inspecting Figure 2.3 and

approximating the mean of the three subjects at the points where carrier frequency $f = 1800$ Hz.

The above-described phenomenon and associated Equation (2.4) was termed the "first effect of pitch shift" by the authors of this study, having been the first to document it in a comprehensive way, though de Boer (1956) had also pointed it out with more limited data. The first effect characterized by Equation (2.4) is thus also sometimes referred to as de Boer's rule (Cariani and Delgutte, 1996). The other obvious experimental manipulation which they also studied was to change the modulation frequency g instead of the carrier frequency f (cf. Equation (2.1)). In doing this, the middle frequency of the complex will remain constant and the frequency spacing will increase such that the bottom frequency gets lower and the top frequency gets higher by the same amounts. Since the frequency spacing is increasing under this manipulation, and a greater frequency spacing indicates a higher fundamental in the case of harmonic stimuli, one might presume the effect of this manipulation would be to raise the pitch. In fact their data show the opposite; the pitch gets slightly lower. This was deemed the "second effect of pitch shift". Just as with the first effect and f, the perception of the second effect can only be maintained within some relatively narrow bounds of g. In general the perception of residue pitch is called "synthetic", and the perception of the components as separate entities is called "analytic". As f for the first effect, or g for the second effect, increases or decreases sufficiently,

synthetic perception breaks down and gives way to analytic perception.

The authors note another subtlety in their data which is not visible in the reproduced Figure 2.3, which is that the pitch shift is slightly and consistently greater than the first effect, Equation (2.4), predicts. They put this also under the category of the second effect, and alter their model for a final form that accounts for both phenomena of the second effect,

$$\Delta p = \frac{(1+b)\Delta f}{n} - b\Delta g, \tag{2.5}$$

where b is a constant. The authors note that b will not be the same for different subjects, and may vary even within a subject for different carrier frequencies; one consistent b they observed for one subject was 0.27. It can be seen from the equation that if there is only a change to the carrier frequency f, the second term is zero and we are left with a modified form of the first effect, Equation (2.4), accounting for the slightly larger than expected pitch shift. If only the modulation frequency g is changed, then the first term is zero and it can be seen that the second term has a slightly negative effect on the pitch, which was what was observed. And if both Δf and Δg are nonzero, then both terms contribute to the pitch shift, although this condition was not tested experimentally.

These experimental results were further interpreted by Schroeder (1966) with a

slightly different model derived from including a term for frequency modulation in the original stimulus equation, thus

$$s(t) = \Big(1 + \cos(2\pi g t)\Big) \sin\Big(2\pi f t - \cos(2\pi g t)\Big) \tag{2.6}$$

where the frequency modulation of the carrier is the same frequency as the amplitude modulation, g Hz. He supposes that this frequency modulation is something the auditory system is doing, however no further physiological motivation is offered. The model's predictions fit the data very well though, and are the lines plotted against the Schouten et al. (1962) data in Figure 2.3. The derived model looks very similar to Equation (2.5), although it is written as a solution for the pitch p instead of the pitch shift Δp:

$$p = \frac{f - g}{n - 1} \tag{2.7}$$

It can be seen that this model has all the necessary effects: It scales the pitch shift inversely with n, the pitch is bigger than expected by the first effect (Equation (2.4)) since we are dividing by a smaller number, and a bigger modulation frequency g results in a smaller pitch. Including no constant to account for individual differences, he seems to assume that the variance in b in the data is due only to noise, certainly not impossible given the lack of power in the study. If one solves for b of Equation (2.5) to obtain the

same pitch shift as given by Schroeder's Equation (2.7), one finds that it is positive, and decreases for increasing n, consistent with Schouten et al. (1962) when they point out that it changed as a function of carrier frequency f.

The interpretations described thus far take for granted that stimuli consist of complexes generated by full amplitude modulation of a sinusoid, resulting in a three-tone complex, where n is the harmonic number of the middle component. In general however we can imagine an arbitrary number of harmonics in the stimuli, where the language of amplitude modulation would no longer apply and n would have to be more well-defined. For instance, should it refer to the harmonic number of the middle component, falling halfway between integers for an even number of harmonics, or should it refer to the second-lowest harmonic number in the complex, or second-highest? Patterson (1973) clarified some of these issues with another experiment, this time using six- and twelve-component shifted stimuli.

In addition to using standards with more components, Patterson (1973) also measured residue pitch perception in a different, but overlapping spectral range compared to Schouten et al. (1962). In terms of the harmonic number of the lowest component of the standard stimuli, Patterson (1973) started with 1 and, for the six-component stimuli tested up to 12, and for the twelve-component stimuli up to 8. In this way he was able to test if models such as equations (2.5) and (2.7) hold for stimuli much lower in frequency than

those of Schouten et al. (1962). In terms of the lowest harmonic number n of a standard stimulus, it was found that below $n = 5$, the slope of the pitch as a function of stimulus frequency no longer changes. Indeed it is obvious that Equation (2.7) does not make sense when $n = 1$, though this is surely a possible stimulus. Equations (2.5) and (2.7) also both predict that the pitch shift Δp should be roughly the same as the frequency shift Δf for $n = 1$ and $n = 2$, respectively. The results of Patterson (1973) indicate that this is not the case, and that for lowest harmonic number $n < 5$, $n = 5$ predicts the correct slope of the lines and hence the correct pitch.

The main result of the study, however, is the comparison of the results from the six-component standards to the twelve-component ones. Patterson (1973) found no difference between those conditions, when the lowest harmonic number n was the same for both. If n in, say, the model in Schouten et al. (1962), Equation (2.5), were interpreted to refer to the middle harmonic number for stimuli with more than three harmonics, different pitches would be predicted for the two conditions in Patterson (1973). But his conclusion is simply that the lowest harmonic number n of a complex, regardless of how many harmonics it has, determines the pitch, for $n > 4$. We can interpret the model from Schroeder (1966), Equation (2.7), in the following way: $f - g$ really just refers to the frequency of the lowest harmonic number of the complex, and $n - 1$ just refers to its harmonic number, both of which were true for the stimuli in Schouten et al. (1962) which had $\Delta g = 0$. Under this

interpretation Equation (2.7) holds up well for the data of Patterson (1973) for $n > 4$.

Nonlinear frequency components (NFCs) generated by the auditory system at various levels have often been invoked to explain pitch perception phenomena such as shift of the residue, and before that, missing fundamental perception in general; indeed this was a basis for early spectral theories such as that of Helmholtz. Shift of the residue is one of the best ways of disproving this, as de Boer (1956) discusses at length. Another convenient way of demonstrating that the distortion product at the difference tone was not responsible, at least in the cochlea itself, for missing fundamental perception was shown by Licklider (1956). There he started with a missing fundamental stimulus, and added a narrowband noise to it which was spectrally centered at the fundamental frequency. If the cochlea was directly responsible for the perception through the generation of the frequency, the noise should eliminate the perception since it would eliminate the cochlea's ability to generate it. Licklider (1956) found that the perception remained intact.

But the influence of the theory that NFCs in the auditory system affect pitch shift continued. Smoorenburg (1970) suggested that odd-order NFCs may be responsible for the second effect of pitch shift. Whereas even-order NFCs like the difference tone are likely not playing a role, odd-order components such as $2f_1 - f_2$ and $3f_1 - 2f_2$, where f_1 and f_2 are the lowest and second-lowest frequencies in the stimulus respectively, may be.
Equation (2.7) accounts for the second effect by the numerator using the lowest frequency

in the stimulus, and the denominator using its corresponding harmonic number. But the second effect could also be explained by the NFCs listed above, since their addition by the cochlea would effectively decrease the lowest frequency f and harmonic number n by one or two. This addition to the stimulus would increase the pitch shift Δp whether a model takes n to be the lowest, second lowest, or middle harmonic number. This hypothesis can also be tested with noise bands.

G. A. Moore and B. C. J. Moore (2003) performed this test by including noise bands at both the region of the even-order difference tone and the odd-order NFCs just below the stimulus frequencies, with multiple fundamental conditions and three resolvability conditions. This experiment also used a larger number of harmonics for the stimuli, to expand the understanding gained from older studies and corresponding models that assumed three-component complexes (Schouten et al., 1962; Schroeder, 1966). Their control was the exact same experiment without the noise bands. They found no differences between the two experiments, and in both cases Schroeder's model, Equation (2.7), predicted the pitch shift results, where $f - g$ is interpreted to be the lowest frequency in the complex and $n - 1$ interpreted to be the corresponding harmonic number. Thus they conclude that the second effect exists regardless of the extent to which the cochlea is producing physical odd-order NFCs, since the noise would have destroyed them. A summary of their results from their first experiment are in Figure 2.4.

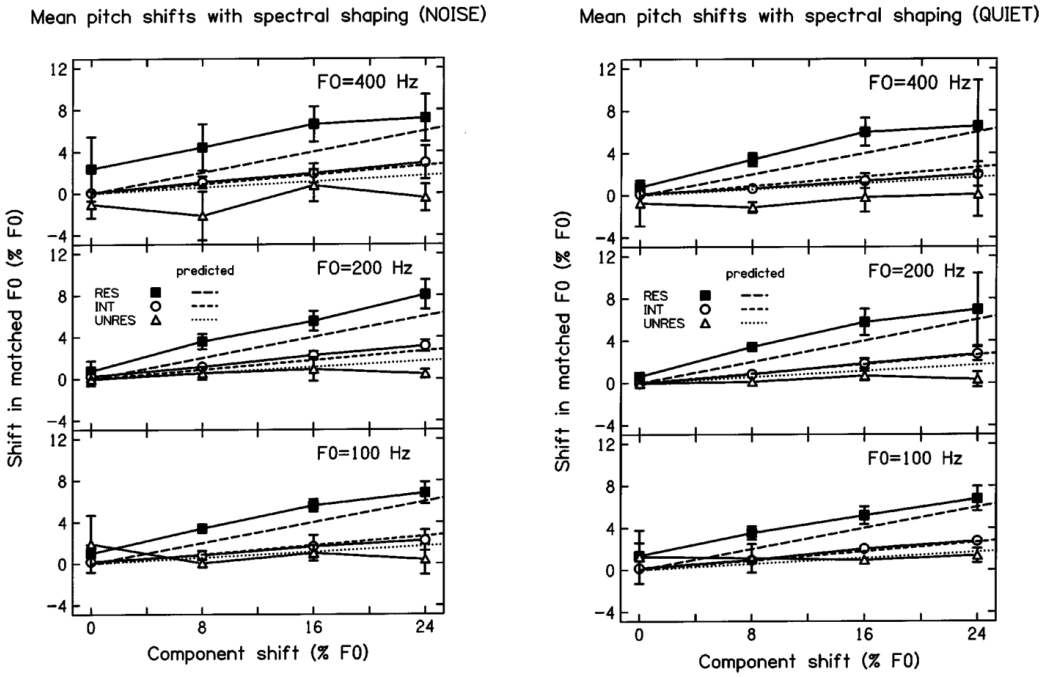


Figure 2.4: Summary plot of pitch shifts comparing the noise condition to the control condition from G. A. Moore and B. C. J. Moore (2003).

In their first experiment, G. A. Moore and B. C. J. Moore (2003) also had a condition in which all harmonics were unresolved, meaning significantly higher than the tenth harmonic. They found no pitch shift for the unresolved condition, and they mention that this could make sense for a variety of auditory theories; if the individual frequencies are not resolved by the auditory system, it would seem that the only information available is the envelope frequency, which as we know does not change for these pitch-shifted stimuli. They also note that this result rather conflicts with the results in Schouten et al. (1962), in

which some stimuli included only unresolved harmonics and where a pitch shift was nevertheless observed. In a second experiment, G. A. Moore and B. C. J. Moore (2003) successfully replicated the relevant results from Schouten et al. (1962). They invoke the notion of "excitation pattern" to explain the differing results of their two experiments. This refers to a general spectral pattern, or in simple cases, just a spectral centroid. They suggest that, rather than matching residue pitches for the unresolved conditions, their results and those of Schouten et al. (1962) show that the subjects were matching excitation patterns, which are in fact very similar for an unresolved shifted standard and the expected matching comparison predicted under either Equation (2.5) or (2.7). In contrast, they controlled for this in their first experiment: For each combination of fundamental and resolvability condition, there was an unchanging spectral passband for all standards and comparisons while shifting. Thus the excitation patterns between standard and comparison were always more or less the same in the intermediate and unresolved conditions.

Their resolved condition was treated in a slightly different way however. Since all frequency components in this condition were surely resolved by the auditory system, subjects may attempt to match specific individual frequencies instead of trying to hear out the residues and match those. To ensure this could not be the case, the spectral passbands used for the resolved condition differed between the standards and comparisons. However for all standards in this condition, the passband was held constant, just as was the case for

the intermediate and unresolved conditions, and similarly for all the comparisons. So it was still the case for the resolved condition that the subjects would not be able to match excitation patterns, since they are always very different between the standards and comparisons. Namely, the spectral passband for resolved standards was centered on the fifth harmonic, and for the corresponding comparisons the passband was centered between the 11th and 12th harmonics. An interesting consequence of this aspect of the resolved condition is that, for some of the fundamental conditions, the residue pitch matches are translated linearly up from the prediction lines based on Equation (2.7) by in some cases many Hz, such that for instance the pitch match to the harmonic situation is higher than the fundamental. This can most clearly be seen in the top left plot of Figure 2.4.

G. A. Moore and B. C. J. Moore (2003) note that, for comparing the noise and control conditions, and indeed for testing a pitch shift model (as opposed to a pitch model), the slopes of these lines are what are important, rather than their intercepts. But they do offer an explanation for the translation, coming originally from Walliser (1969), who points out that a sinusoid has a consistently higher pitch than a stack of its own harmonics. In the resolved condition in G. A. Moore and B. C. J. Moore (2003), the comparison stimuli had more harmonics, and higher harmonics, than the standards, so this may be an example of a generalization of the principle pointed out by Walliser (1969). Even for the harmonic situations, the comparison stimulus with a fundamental properly at

the non-shifted fundamental of the condition would have a lower pitch than the standard according to this principle, thus the subjects would have shifted the comparison up a little to compensate. This is clear in the results in Figure 2.4, and is fairly consistent through the different levels of shift.

2.4 Concluding remarks

Though there is a fair amount of variability in the existing data, presumably because of the difficulty of the task and subtle differences between stimuli, it is clear that the shift of the residue is a very lawlike phenomenon. Whatever one's precise model, many authors have pointed out since de Boer (1956) that the residue shift is consistent with the inter-peak intervals of the shifted time-domain waveform. As is observable in Figure 2.2B compared to 2.2A, the period of time between successive peaks is slightly less for the complex which is shifted up, resulting in a slightly higher pitch frequency. For complexes with higher harmonics there are more time-domain peaks that are closer together in amplitude, and Schouten et al. (1962) and others point out that the different possible peaks to choose from correspond to the multiple pitches that are indeed perceived for a given complex, depending the perceptual context.

In general it seems, then, that autocorrelation and related transformations may be sufficient to explain much of the phenomenon. Indeed models based on autocorrelation of stimulus waveforms or some early auditory correlate such as auditory nerve action potentials have proven to be able to explain a wide variety of pitch phenomena (Cariani and Delgutte, 1996). One problem with simple explanations based on autocorrelation is that there do not seem to be any neurophysiological structures capable of delaying an auditory signal for the times that would be necessary for what we know about pitch perception (de Cheveigné and Pressnitzer, 2006). Another problem is evident when looking at the plotted autocorrelation function of the shifted complex in Figure 2.2B. A simple model based on autocorrelation would take the highest peak as the reciprocal of the predicted frequency, but while a peak for the correct pitch perception does exist in this function (≈4.6 milliseconds), it is not the highest peak. The highest peak is at 20 milliseconds, which corresponds to 50 Hz. And indeed the three frequencies in this shifted complex are all harmonics of 50 Hz, though not consecutive harmonics. 50 Hz is a fairly low frequency from the standpoint of the human auditory system but is certainly within the limits of perception. Nevertheless this shifted complex is much more likely to generate the shifted residue pitch of ≈216.6667 Hz than the lower residue pitch which is actually a common subharmonic. We can point out that, strictly speaking, most of these shifted complexes are not truly "inharmonic" even though they are commonly referred to that way.

The only truly inharmonic complexes would contain frequencies at irrational ratios to one another. A more constrained and proper definition of an inharmonic complex in this case would simply be a complex without consecutive harmonics of any fundamental frequency.

Theories of pitch perception, including higher level music cognition, often invoke common, or nearly-common subharmonics of concurrent frequencies to explain many perceptual phenomena (Terhardt, 1974). Schroeder (1966) pointed out that such an explanation also works nicely for residue shift phenomena. In the case of Figure 2.2B, the pitch frequency should be $650/3 = 216.6667$ Hz according to Schroeder (1966) and confirmed by Patterson (1973) and G. A. Moore and B. C. J. Moore (2003), which is of course an exact subharmonic of 650, the first component frequency, and is a near-subharmonic of the other two frequencies. Schroeder (1966) also points out however that this model makes nearly identical predictions to a delay-based or autocorrelation-based model.

Phenomena of the residue, particularly pitch shift, provide a unique window into the workings of the auditory system and whatever subset of it is responsible for pitch extraction. Thus with even the simple experimental manipulation of creative stimuli, we can probe the perceptual apparatus fairly deeply. Questions of course arise as to the underlying physiology of these perceptual phenomena. One thing that makes delay-based models broadly attractive is because of the increasing awareness that the auditory system

does not physically generate components at pitch frequencies in general. While the auditory system generates a wide variety of NFCs, both even- and odd-order, the pitch of the shifted residue is not among them (Gockel et al., 2011). The even-order difference tone is however among them, so the pitch of a simple harmonic missing fundamental stimulus could be mistakenly attributed to this NFC, and indeed has been. Given these impasses, the psychophysics and physiology behind residue pitch perception will likely remain an important tool for studying the auditory system in the future.

Chapter 3

A high-density EEG FFR source analysis study

3.1 Introduction

The frequency following response is a well-characterized and useful tool for studying auditory function in both research and clinical settings (Skoe and Kraus, 2010). However a number of its aspects remain elusive or unexplained. For instance, while researchers generally expect certain nonlinearities of the types mentioned in Chapter 1 to appear in FFRs, a precise account of exactly which frequencies should appear as a function of the contents of a stimulus has yet to be done. And perhaps more urgently, there is still little

consensus about the neural and peripheral generators of the FFR. Therefore a novel and exploratory experimental paradigm was conceived to address these gaps in the literature.

First, careful attention was paid to stimulus design. To learn which nonlinear frequency components are added by the brain and periphery, it is necessary to use inharmonic stimuli. This is because, as discussed in Chapter 2, in the case of harmonic stimuli, the odd-order nonlinearities overlap with the even-order frequencies. Thus, the even- and odd-order portions of the response are composed of the same frequencies. As an example, a synthesized speech syllable is often used to elicit FFRs because of its reliability and robust response (Skoe, Burakiewicz, et al., 2017; Skoe and Kraus, 2010). As is the case with all real or synthesized human speech, this "da" has a harmonic spectrum. With this and similar stimuli, researchers find that the FFR contains the same spectral frequencies as the stimulus, with the relative amplitudes of the frequencies in the spectrum weighted differently; lower frequencies, and the fundamental in particular, are weighted most heavily. So it is clear that the brain is transforming the stimulus in some way, but it is not clear whether the brain is adding any frequency.

Using simpler and carefully-controlled stimuli, it is readily possible to ascertain which frequencies are being generated by the brain and cochlea, and also to tease apart even- from odd-order components. In fact, the shifted missing fundamental stimuli described above are ideal examples. Thus the stimuli in the present study were modeled after Gockel et al.

(2011), who utilized shifted missing fundamental sounds. The amount of frequency shift in such stimuli can be expressed as a ratio of the shift to the fundamental. It will be recalled from Chapter 2 that, in the case of this type of stimulus, the QDT, envelope frequency, and difference frequency are also this frequency. The stimuli in Gockel et al. (2011) were shifted by ratios of $\frac{1}{2}$ and $\frac{1}{4}$, but to be sure which frequencies in the FFR were related to the "envelope" of the stimulus, the present study utilized irrational shift ratios (Table 3.1).

The neural and peripheral generators of the FFR have long been a topic of interest. Much of this research has started from the assumption that this response is fully generated in subcortical auditory structures, and ventured to determine the relative weightings from those structures. Sohmer et al. (1977) utilized clinical populations with lesions or hearing loss and determined the main source of the FFR to be the inferior colliculus (IC), while Gardi et al. (1979) used invasive approaches in cats and concluded that the dominant source was the cochlear nuclei (CN), with lesser contributions from the IC and the cochlear microphonic.

There have also been more recent efforts utilizing modern neuroimaging techniques. Bidelman (2015) used a 64-channel EEG paradigm and determined the main source of the FFR was the IC. However Coffey et al. (2016) performed a high-density MEG study to localize the FFR to a speech syllable sound and found, in addition to sources in the auditory brainstem and thalamus, a bilateral cortical contribution from Heschl's gyrus.

Compared to EEG, MEG is more sensitive to superficial sources and tangential source dipoles in general, while EEG is sensitive to both radial and tangential sources and is comparatively more sensitive to deep sources (Ahlfors et al., 2010; Goldenholz et al., 2008). Thus, it is possible that MEG overemphasizes the cortical result. Additionally, because these results contradict previously-held assumptions, it is important to attempt to replicate them using EEG. Therefore an experiment was undertaken 1) to specify which nonlinearities are generated in the FFR and 2) to be the first high-density EEG FFR source analysis study.

3.2 Methods

3.2.1 Summary

This study is meant in part to replicate the results of Coffey et al. (2016), but also to attempt to shed some new light on the FFR by precisely characterizing nonlinear components in the FFR and finding their sources. Not only is it the first FFR study to attempt source localization with EEG and individual anatomical data, but it also utilizes novel stimuli to ask whether it is appropriate to refer to the even-order portion of the response as "envelope-following". To accomplish this, the study's data acquisition first involved a ten-minute T1-weighted anatomical MRI scan. Next, participants were fitted

with a 256-electrode net manufactured by EGI, using an EEG amplifier capable of a high sampling rate appropriate for FFRs. Before EEG acquisition, twelve pictures were taken of each participant's head with the net firmly fitted using EGI's hardware and software. These pictures aid in localizing the electrodes along with three fiducial points for co-registration with the MRI scan. EEG data then comprised roughly 60 gigabytes per participant, thus processing was done on a computing cluster specialized for high-memory computational needs. Once EEG data was downsampled and trials averaged, the remainder of the analysis and source localization was done on personal computers.

3.2.2 Participants

After a single pilot participant was run to test the functionality of the acquisition and data analysis pipelines, twelve participants were recruited for this experiment. Participants were recruited based on the fact that all had previously undergone FFR acquisition in unrelated studies and were known to have a robust FFR. Having a relatively low signal-to-noise ratio in an FFR to complex sounds does not indicate any problem of processing or function, nor does having a higher signal-to-noise ratio necessarily indicate greater perceptual abilities. However for this study's purposes, better FFRs were preferred to more accurately characterize the various responses, as desired. The average age of the participants was 23 ($SD = 2.76$) and all were female. All participants had normal hearing thresholds as had

been previously established from audiometric measures when they participated in earlier FFR studies. Participants were monetarily compensated at the rate of $25 per hour, including the pilot participant. All experimental subjects received exactly $50 as the entire experiment was very nearly two hours each time. Informed consent was obtained from all participants and the study design was approved by the University of Connecticut Institutional Review Board.

3.2.3 General study design

These participants had never been exposed to these particular stimuli before, nor had any of them ever had a structural MRI scan. All data acquisition including EEG and MRI for each subject was done in one block of time so efficiency of experimental structure was important. All data was acquired at the University of Connecticut Brain Imaging Research Center (BIRC). Three hours were blocked out for each participant, however after efficiency was achieved, only two hours were required. Upon arriving, participants first read consent documents and signed them, and also filled out a brief questionnaire about music and language background. They also filled out an MRI safety screening form. Immediately upon finishing these documents, their head circumference was measured so the appropriate EGI EEG acquisition net could begin soaking in a potassium chloride electrolytic solution while the participant's MRI was being acquired. After head measurement, participants

were led to the MRI room where the BIRC MRI technician checked their MRI safety screening form and performed the MRI acquisition. While the primary investigator accompanied each participant to and from the MRI, an assistant began soaking the EEG net in electrolyte so that it was ready to be applied upon completion of the MRI scan.

Participants were then led to an EEG acquisition room. Outside the sound booth, the soaked EEG net was put on their head and fastened. They were then led to another room that housed the EGI photogrammetry hardware and software, and twelve pictures were acquired from different angles to co-register the EEG data with the MRI data for source analysis. After photograph acquisition, participants were led back to the EEG room and into the sound booth where EEG acquisition would take place. The EEG net was prepared for acquisition, and participants were told how the remainder of the study would unfold. Binaural ear inserts were then placed in their ears to deliver the stimuli and they began watching a silent movie. No behavior of any kind was elicited or measured from the participants, other than a request to minimize eye blinks and body movements during perception of the stimuli. The sound booth door was closed, and stimulus delivery and EEG acquisition began. EEG acquisition was approximately 51 minutes, including 20-second rest periods between each of eight blocks of stimulus delivery, but there were otherwise no breaks. After data collection was finished, the participants were led out of the sound booth and were helped out of the EEG net. They were then debriefed and

compensated.

3.2.4 Stimuli

The stimuli for this study were designed to separate even- from odd-order components of the FFR, and specifically to ask whether frequencies not related to the stimulus envelope were present in the even-order portion, often simply referred to as the "envelope-following response" (Aiken and Picton, 2006; Aiken and Picton, 2008; Dolphin and Mountain, 1992). Even- and odd-order portions of the FFR are obtained by delivering stimuli in two opposite polarities, averaging those two groups of trials separately, and then summing the two polarity conditions to obtain the even-order response, and subtracting them to obtain the odd-order response. Historically, many FFR experiments have utilized the alternating polarity technique to avoid electromagnetic stimulus contamination in the electrodes. One can indeed be assured that no stimulus artifact is present in the even-order portion of the response, since stimulus artifact is by definition first-, and therefore odd-order. However as mentioned above, if only the even-order response (which includes the envelope) is analyzed, the entire odd-order portion, which contains responses from the brain and periphery to the stimulus primaries themselves, as well as other higher-order odd-order responses such as the well-known cubic difference tone (CDT), is not considered. Thus stimuli were presented for the present study in both polarities in order to analyze the content of both the even-

and odd-order portions of the response.

The even-order portion of the FFR is often associated with the envelope of the stimulus because it indeed often contains a prominent component at the amplitude envelope frequency of the stimulus, as well as some harmonics. In a harmonic stimulus such as a speech syllable, the envelope frequency is also the frequency difference between successive harmonics, so even if there is little or no energy at the fundamental frequency in the stimulus, the auditory system produces prominent energy there. It is in fact these nonlinear relationships between stimulus components, such as difference or summation, that predicts their presence in the brain's response (Lerud et al., 2014), rather than just the amplitude envelope of the stimulus.

If one utilizes shifted missing fundamental stimuli of the type described above, the auditory system's addition of the difference frequency should remain unchanged from the harmonic case, but if the even-order portion is generating nonlinearities beyond this, they would not necessarily be predicted to be harmonics of the envelope frequency. For instance, a stimulus consisting of 205 and 305 Hz can be imagined as a shifted missing fundamental stimulus with a nominal fundamental of 100 Hz. The brain's response to this would include the difference frequency of 100 Hz, produced as the difference of the two primaries. However the sum of the primaries is 510 Hz, no longer a harmonic of the "envelope". The question as to whether or not the FFR includes nonlinearities of this type has not been

	Primary 1	Primary 2	Primary 3	Shift ratio	QDT	CDT
Stimulus 1	178.8562	258.8562	338.8562	$\frac{\sqrt{2}}{6}$	80	98.8562
Stimulus 2	174.1421	254.1421	334.1421	$\frac{\sqrt{2}}{8}$	80	94.1421
Stimulus 3	469.4975	679.4975	889.4975	$\frac{\sqrt{2}}{6}$	210	259.4975
Stimulus 4	457.1231	667.1231	877.1231	$\frac{\sqrt{2}}{8}$	210	247.1231

Table 3.1: A table specifying the parameters of the four stimuli used in the present study. Irrational numbers are approximate. Frequencies are in units of Hertz. Stimulus primaries are harmonics 2, 3, and 4 of a missing fundamental specified as the QDT, shifted up by the specified ratio. For all these stimuli, the envelope frequency is equal to the QDT.

sufficiently explored, and inharmonic stimuli can be of use in this regard. Gockel et al. (2011) also used pitch-shifted stimuli, but the ratios of shift were only $\frac{1}{2}$ and $\frac{1}{4}$. One would indeed expect only harmonics of the envelope in the $\frac{1}{2}$ shift condition with the above calculations in mind. However in the $\frac{1}{4}$ shift condition, that study found some FFR frequencies that were not harmonics of the envelope, namely harmonics of the first subharmonic of the fundamental. Keeping with our simple example above, we imagine a stimulus consisting of 225 and 325 Hz. The summation frequency of 550 Hz is not a harmonic of 100 Hz, but it is a harmonic of 50 Hz. If the brain were producing 550 Hz because it is a harmonic of 50 Hz, one might still consider it "envelope-related". By using irrational ratios, the present stimuli help to more clearly investigate which nonlinear process produces which frequencies.

Four different pitch-shifted stimuli were used, summarized in Table 3.1. Each stimulus was a three-component complex consisting of nominal harmonics 2, 3, and 4 of a missing

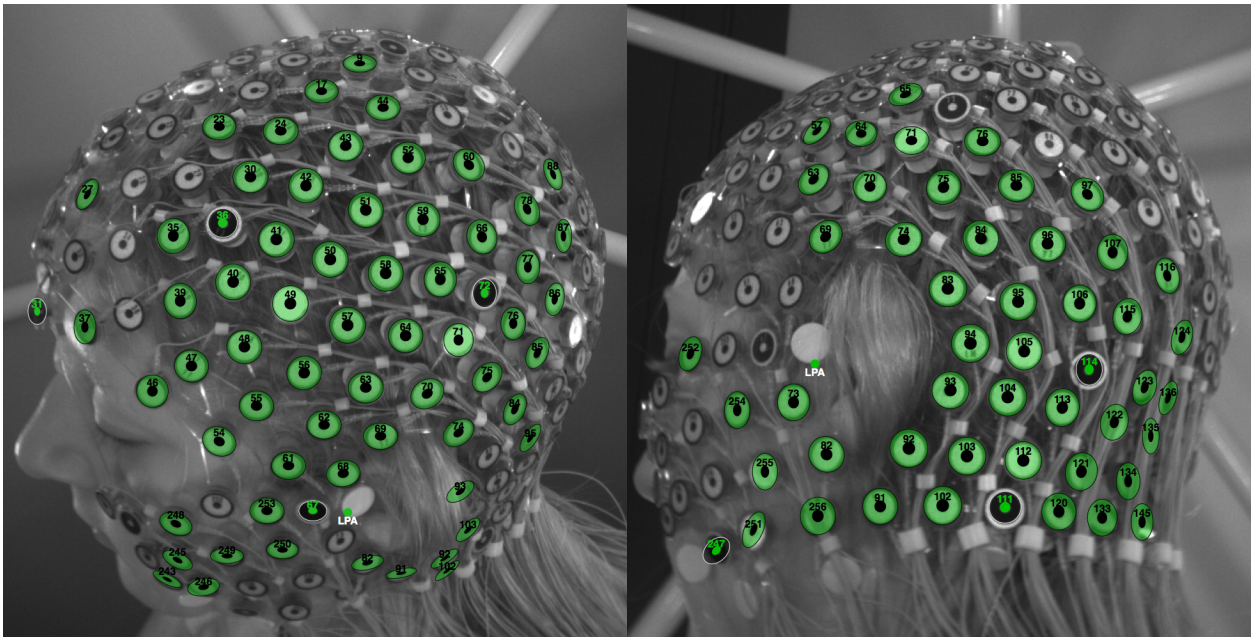
fundamental, shifted up by an irrational ratio. Some expected nonlinearities are specified in the Table as well. Stimuli contained linear on and off ramps of 5 milliseconds each. Each trial was 350 milliseconds long, followed by a random inter-stimulus interval with a mean of 137.5 milliseconds and a possible spread of 25 milliseconds in both the negative and positive directions. Inter-stimulus intervals came from a uniform distribution. The four stimuli were delivered in both polarities, thus the experiment consisted of eight unique stimuli. These eight stimuli were presented 750 times each, in a pseudorandom order distributed throughout eight blocks. Thus each block was approximately six minutes, and there was also a 20-second break between each block. The random parameters in the stimulus generation were intended to minimize any habituation that might occur due to short-term plasticity, thus theoretically maximizing the FFR's signal-to-noise ratio. The stimuli were delivered in the same order for each participant, and this order was known a priori so that responses could be grouped accordingly during data analysis. The stimuli were calibrated to be 70 dB SPL at the eardrum.

The stimuli were generated as `.wav` files in MATLAB (The Mathworks Inc., MA, USA). The right channels of the files contained the stimuli and were split and delivered to both ears, while the left channel contained a DC trigger event marker that was delivered along with the EEG data to ensure timing accuracy. The stimuli were delivered with an Etymotic transducer and insert earphones with foam tips. The transducer was

mu-metal-shielded to minimize stimulus artifact contamination, and hung approximately one half meter below the EEG electrodes along the participant's torso.

3.2.5 EEG and MRI acquisition and processing

EEG data was amplified with an MR-compatible EGI NetAmps 410 amplifier. The purpose of using this amplifier was not to collect data in-scanner, but was that it was capable of a higher output sampling rate than the more typical NetAmps 400. Thus EEG data was recorded at the 410's maximum rate of 20,000 Hz, a much more desirable rate for FFR acquisition than the maximum rate of 1,000 Hz for the NetAmps 400. EEG was recorded using EGI's Net Station software. Participants watched a silent movie on a screen approximately one meter in front of their heads while data was being collected. They were instructed to minimize eye and body movements while they were hearing the auditory stimuli, but were otherwise not instructed to behave or react to anything. Once the EEG net was fitted to their head, they were taken to a room with an EGI Geodesic Photogrammetry System (GPS) and pictures were taken for electrode localization with EGI's GPS Solver software. An example of what the fitted EGI EEG net looks like within the solver software is provided in Figure 3.1. Before acquisition, the net preparation was finished by ensuring that all electrode sponges were in as close contact as possible with the scalp. More electrolytic solution was individually applied to any sponges that required it.



(a) View from camera 6. (b) View from camera 10.

Figure 3.1: Two views of the fitted EGI Geodesic net from EGI GPS Solver electrode localization software.

Electrode impedances were kept below 50 kΩ, which is typical for a high-impedance system such as EGI's.

EEG processing was done in MATLAB using the free and open source (FOSS) EEG/MEG analysis software package FieldTrip (Oostenveld, Fries, et al., 2011). At the initial sampling rate of 20,000 Hz, each block of EEG was approximately 15 gigabytes in memory when loaded in MATLAB, which is prohibitive on a normal computer. Thus UConn's High Performance Computing (HPC) cluster was utilized for EEG processing in FieldTrip. Upon initial data analysis, an electrical artifact at 1,000 Hz and its harmonics

was discovered which contaminated all blocks in all subjects. Because brain data beyond this frequency could not be easily analyzed if present, and because the initial rate of 20,000 Hz is unwieldy and greater than necessary, subsequent analysis resampled all EEG to 2,500 Hz. After resampling, data was subjected to a frequency-domain bandstop filter at the electrical supply AC frequency of 60 Hz and its harmonics. The main filter was a bandpass between 63 Hz and 950 Hz. The low cutoff was arrived at through visual inspection of the filter's magnitude response to be the highest possible frequency that also does not affect the lowest frequency expected in the EEG responses, namely the lower QDT of 80 Hz. The high cutoff was the highest frequency that also completely filtered out the electrical artifact at 1,000 Hz. This was a linear-phase, 500th-order, finite impulse response filter, implemented as zero-phase with MATLAB's `filtfilt()` function. This filter was applied to the continuous EEG data before epoching, and also had the convenient effect of making all the data zero-mean, which is expected from any sufficiently aggressive highpass filter.

Also before epoching, an automated channel repair procedure was applied to each block of the EEG data. A FASTER-like (Nolan et al., 2010) algorithm selected statistically-outlying channels that needed repair, and FieldTrip's own channel repair function was then utilized. This function replaces the data in each of the outlying channels with a spline interpolation of the data from neighboring channels. The distribution of the number of channels that were repaired in each block, for all participants, failed a

one-sample Kolomorov-Smirnov test of normality, hence non-parametric descriptive statistics are appropriate ($Mdn = 6.5$, median absolute deviation (MAD) $= 1.5$).

The continuous EEG data was then epoched according to the event times that were recorded. There were eight blocks for each subject consisting of 750 trials each. There were four different stimuli, each with two polarity conditions, and these eight unique stimuli were distributed pseudorandomly throughout the eight blocks. After epoching, all EEG channels were re-referenced to a global average for the purpose of later source analysis. Two methods of trial rejection were then applied. The first was another FASTER-like algorithm to detect statistically outlying trials. The second rejected trials based on a threshold of absolute amplitude, under the assumption that trials with very high amplitude were likely contaminated with eye or muscle artifact. The threshold for the latter routine was 65 μV. The distributions for the number of rejected trials for both of these methods were also non-normal. The FASTER method ($Mdn = 13$, $MAD = 4$) typically rejected slightly more than the absolute amplitude method ($Mdn = 11$, $MAD = 9$). Once artifact rejection was complete, trials were then grouped according to which stimulus they corresponded to, and ERPs for each stimulus were generated by averaging over trials. For each stimulus, the even-order portion of the response was then created as the sum of the responses to the two polarity conditions, and the odd-order portion was created as the difference between them.

In order to construct accurate and individualized forward models for source analysis,

each participant underwent an anatomical T1-weighted structural MRI scan before EEG collection. The MRIs took approximately 10 minutes and were collected on a Siemens Prisma 3 Tesla scanner located in the BIRC. The FOSS package FreeSurfer (Fischl, 2012; Fischl et al., 2002) was used to segment the entire brain from the raw MRI image. Neocortex was reconstructed as a surface of vertices, each with an orientation, and subcortical structures were also parcellated as a volume of vertices without orientations. The results of FreeSurfer's reconstructions were then imported into the FOSS brain imaging package Brainstorm (Tadel et al., 2011), along with the ERP matrices for each stimulus, for source analysis.

3.2.6 EEG source analysis of the FFR

After import into Brainstorm, an affine transformation matrix to the MNI152 template brain was created for each FreeSurfer reconstruction. This transformation allows for the transfer of areas of interest in the brain between participants' brains, with the MNI152 template brain as an intermediary. It also allows for automatic identification of six fiducial points required to successfully co-register the MRI with the EEG electrode locations. These points are the nasion, left and right preauricular points, anterior commissure, posterior commissure, and an interhemispheric point (located mid-sagittally) in the upper portion of the head volume. The nasion and both preauricular points were also identified in

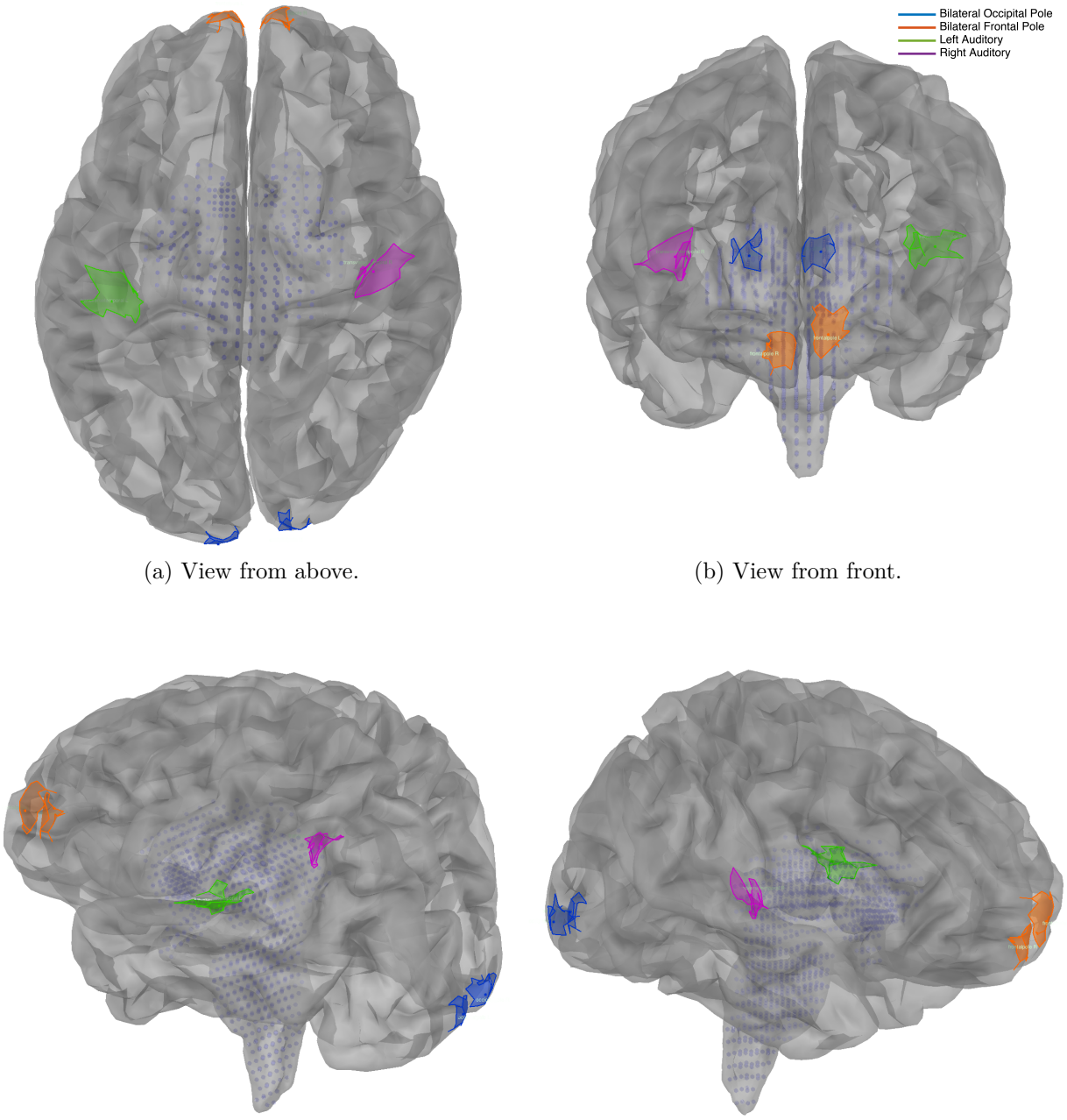
the EGI GPS Solver software along with the electrodes; thus the electrode locations can be superimposed on the MRI reconstruction for each participant, allowing for accurate individual forward modeling.

A mixed source model, which contains both surface (orientation-constrained) and volume (not orientation-constrained) vertices, was created in Brainstorm for each subject's head. The entire brain except the cerebellum was used for each source model, containing approximately 24,000 vertices per brain. A forward model was then calculated for each subject using the Boundary Element Method as implemented in the FOSS package OpenMEEG (Gramfort, Papadopoulo, et al., 2010; Gramfort, Papadopoulo, et al., 2011) which works within Brainstorm. A forward model is a transformation that gives a scalp-space representation of time series data given the source-space data. Typically, however, researchers want a model that does the opposite: Given the scalp-space time series data, a transformation that yields the source-space data is desired. This approximation is called an inverse model, and the forward model is required as a first step to obtaining it.

For the source computation with an inverse model to be meaningful, several regions of interest (ROIs) within the source model were established. An ROI, or scout, consists of several vertices whose time series will be combined once an inverse model is applied. The scouts were chosen here to be substantially similar to those of Coffey et al. (2016), in an attempt to replicate the finding of cortical contributions to the FFR from primary auditory

cortex, but not from control cortical regions, with the primary producer of the FFR being from several subcortical auditory structures. Thus four cortical scouts and three subcortical scouts were selected. The cortical scouts were: left and right primary auditory cortex, bilateral frontal pole, and bilateral occipital pole. The subcortical were: bilateral cochlear nucleus (CN), bilateral inferior colliculus (IC), and bilateral medial geniculate body (MGB).

An atlas is a grouping of many labeled scouts for organized source analysis, and several atlases come with FreeSurfer's default reconstruction parameters. The frontal pole scout was taken from the Desikan-Killiany atlas (Desikan et al., 2006), as were both primary auditory scouts, labeled by identifying the transverse temporal (Heschl's) gyrus in both hemispheres. The occipital pole scout was created manually. Cortical scouts created for source analysis are depicted in one subject's brain in Figure 3.2. None of the atlases parcel out the brainstem or thalamus beyond identifying them as two different structures, thus subcortical scouts were also created manually. The CN was created by noting the caudal base of the pons, the IC by noting the prominent anatomical features of the corpora quadrigemina, and the MGB by noting the caudal, medial portions of the bilateral thalamus. All seven scouts were made to be exactly 50 vertices by pruning or growing the scout through a nearest-neighbor search. Subcortical scouts created for source analysis are depicted in one subject's brain in Figure 3.3. Auditory and frontal scouts were all identified



(a) View from above.

(b) View from front.

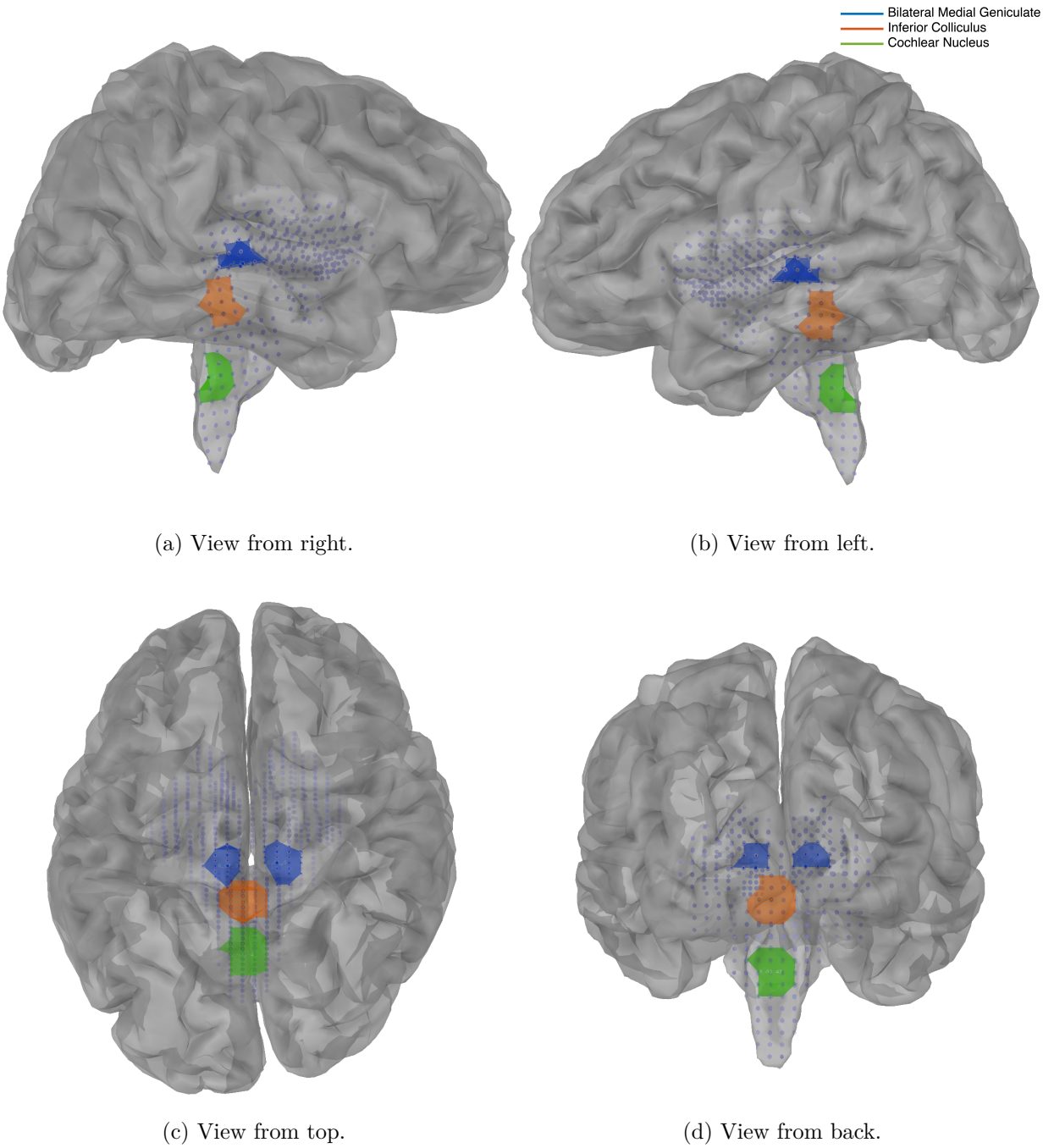
(c) View from left.

(d) View from right.

Figure 3.2: Four views of one participant's brain with cortical scout ROIs marked.

automatically by FreeSurfer for each participant, so there was no transfer between brains necessary. The occipital pole scout was drawn manually in one subject and transferred to the rest of the subjects through each their transformations from the MNI brain. Subcortical scouts were entered by specifying the seed vertices for each one with MNI coordinates that were then transformed to each specific brain, and grown to 50 vertices.

Once the scouts were chosen for source analysis, an inverse model was created for each subject to transform the scalp-space data to time series in source space. A minimum norm estimate (MNE) was used to accomplish this (Gramfort, Luessi, et al., 2014). When the inverse model is calculated, a kernel is created to transform scalp-space EEG data to any desired vertices in the source model. Thus the source-space time series for each scout were calculated and analyzed for their frequency content. Each scout contained 50 vertices, so a single time series was obtained from each scout by averaging the time series of the individual vertices that comprise each scout. The volume (subcortical) vertices do not have an orientation constraint, so Brainstorm gives three time series for each of those instead of one. Each of these represents an orthogonal axis of orientation. Frequency content of volume scouts was thus calculated by averaging each of the three orientation conditions in the spectral domain. All frequency analysis was done as the magnitude of a Fourier transform.



(a) View from right.

(b) View from left.

(c) View from top.

(d) View from back.

Figure 3.3: Four views of one participant's brain with subcortical scout ROIs marked.

3.3 Results

3.3.1 Summary

Scalp-space EEG data was first analyzed to compare it with known aspects of the FFR. Electrodes were averaged over subjects in both the time and spectral domains, in case responses were at different phases for different subjects. Time-domain averages in scalp space were done for visualization, but spectral averages were analyzed for their frequency content, rather than frequency-transforming the time-domain averages. The even-order portion of the responses showed robust amplitudes in many electrodes at predicted frequencies such as the QDT and its first harmonic. Additionally, amplitude was observed at other second-order nonlinearities, namely summation combination tones of the stimulus primaries. This result shows that the even-order portion is not an "envelope-following" response per se, because these summation tones are not harmonics of (or in any way related to) the envelope frequency.

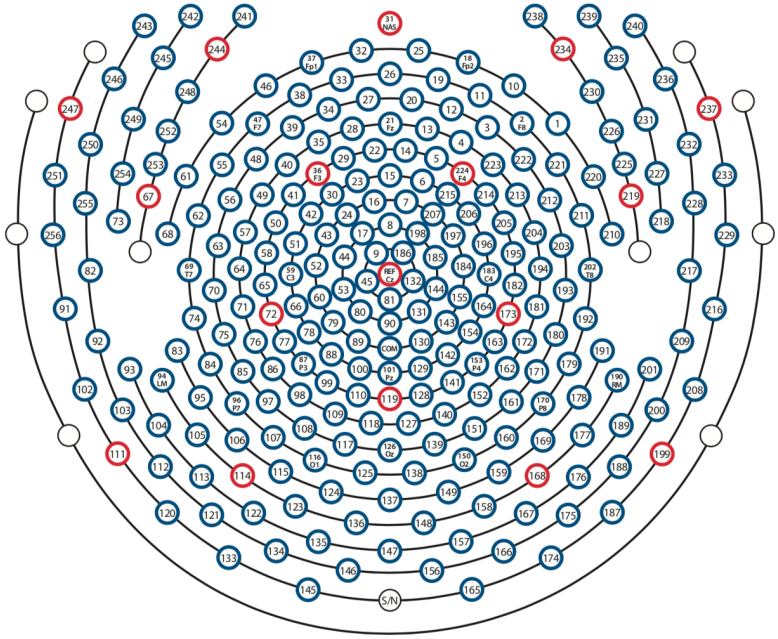
Source analysis was done by creating an inverse model, and then doing a frequency analysis of each of the seven scout time series, for every subject, for every stimulus, for both the even- and odd-order portions of the responses. Scouts were averaged across subjects in the spectral domain. For the lower-frequency stimulus, there was significant amplitude at the QDT from both primary auditory cortices, while control cortical scouts

did not have significant amplitude. Left auditory QDT amplitude was significantly greater than right auditory. No other significant frequencies were found for the lower stimulus in cortex. Frequency analysis of subcortical scouts showed robust responses at multiple FFR frequencies, such as the QDT, its first harmonic, and three second-order summation tones, all of which were significant. The amplitudes of the QDT increased as the auditory system was ascended; thus the MGB had more amplitude than the IC, which had more than the CN. Analysis of odd-order responses also showed a significant CDT in both the subcortical scouts and auditory cortices, while not showing this frequency in the control cortical scouts.

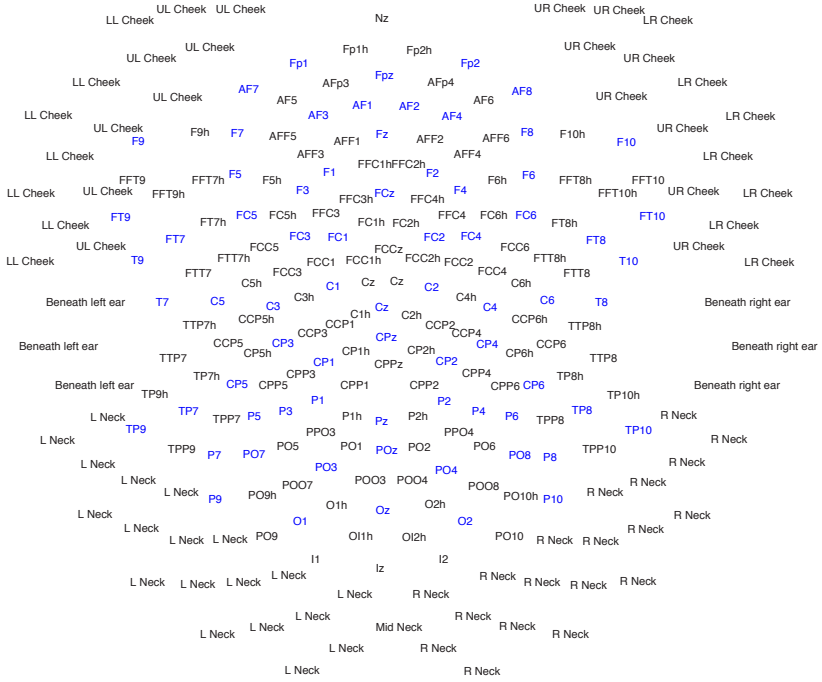
Source analysis for the higher stimulus showed slightly different results than the lower stimulus. For the higher stimulus, there was significant amplitude at the QDT in all four cortical scouts; however, the amplitude of the QDT in both auditory cortices was significantly greater than in both control cortical regions. There were no other significant even-order peaks in the cortical scouts. The subcortical scouts showed robust and significant amplitude at the QDT and its first harmonic. The summation frequencies seen for the lower stimulus were above the filter's highpass cutoff frequency in this case, and so were not observed if they were present. A prominent CDT was found in the odd-order portion of the subcortical scouts, however this frequency was not present in any of the cortex scouts.

Upon initial data analysis, a trigger-related artifact was found during the first 60

milliseconds of the even-order portion of most trials, in most subjects. The intensity of the artifact varied across electrodes. Additionally, there was evidence of stimulus artifact in the odd-order portion, even though mu metal shielding was used around the transducer. These artifacts were not fully resolved during data collection, however two analysis methods were utilized to minimize their effects. The first analysis method to minimize the effect of the trigger artifact was simply not to analyze the first 60 milliseconds. Thus, for source-space data, a Tukey window was utilized that zeroed out the contaminated beginning of each ERP before frequency analysis. In the scalp-space data, a second method of analysis was utilized for data visualization. It was found that a simple PCA captured both artifacts very well in the first two principal components of each ERP matrix. Across all stimuli and subjects, the first two components combined explained an average variance of 72.23% ($SD = 19.22\%$). Thus these first two components were removed from all data. Additionally, there were between three and five samples immediately before and after stimulus onset in the data that were clearly errant and which survived the PCA reduction. For data visualization, those samples were replaced with low-level Gaussian noise in all scalp-space data.



(a) EGI's sensor numbers and locations.



(b) Corresponding 10/5 labels and locations.

Figure 3.4: A novel 10/5 system of sensor location labels for a geodesic sensor net. Sensors in blue were already labeled in EGI's documentation; all others were labeled manually by the author.

3.3.2 Scalp-space FFRs

EGI's 256-electrode net is arranged such that sensors are placed along geodesic curves around the head. This system of sensor placement is different than the commonly-used 10/20 system; however, labels of sensor locations based on 10/20 (e.g. Cz, F3, P2) are both familiar and convenient. Such a system of labels and scalp landmarks for a dense EEG net would use 5% increments along the circumference of the head and would thus be called a 10/5 system, as 10/20 and 10/10 use 20% and 10% increments respectively. A 10/5 system has been proposed (Oostenveld and Praamstra, 2001) and refined (Jurcak et al., 2007), however a transformation from a 10/5 system to EGI's Geodesic Sensor Net has not been done. It is desirable for efficient scientific communication to obtain these sensor location approximations, thus this was done manually prior to scalp-space analysis and is depicted in Figure 3.4.

For scalp-space data visualization and to confirm that a reliable FFR was appearing across subjects, all time-domain electrode data were averaged over subjects, for each stimulus condition. Additionally, spectra were calculated for each electrode separately and averaged across subjects in the spectral domain. To choose four representative electrodes for visualization, a simple measure of SNR was calculated for each electrode's spectrum. The ratio of the QDT peak amplitude and the average noise floor of the summation spectrum was calculated for each electrode, and the highest four were chosen for plotting,

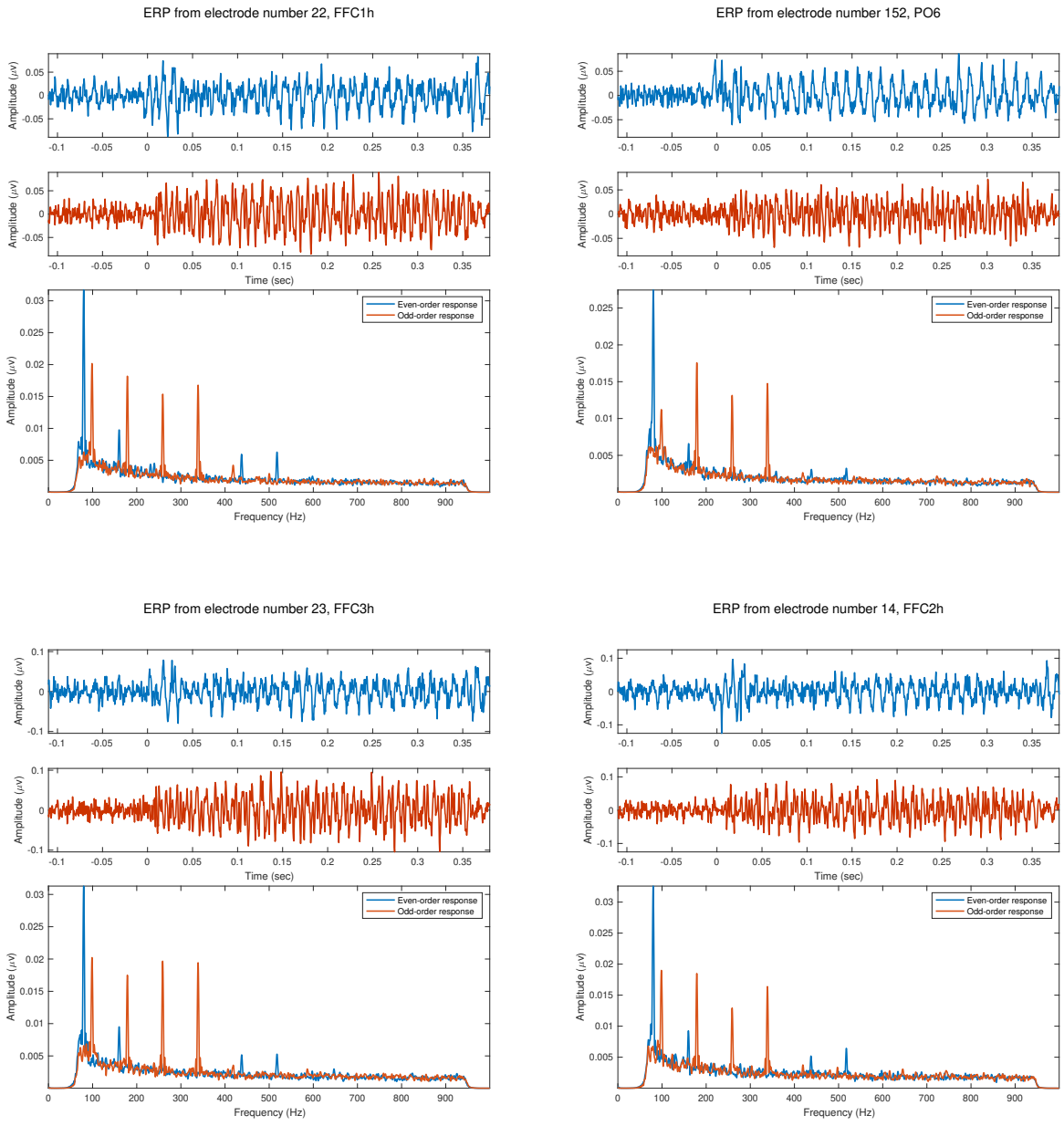


Figure 3.5: Example average scalp-space FFRs for low-f0, large shift stimulus. Time-domain waveforms are averaged in the time domain; spectra are averaged spectra.

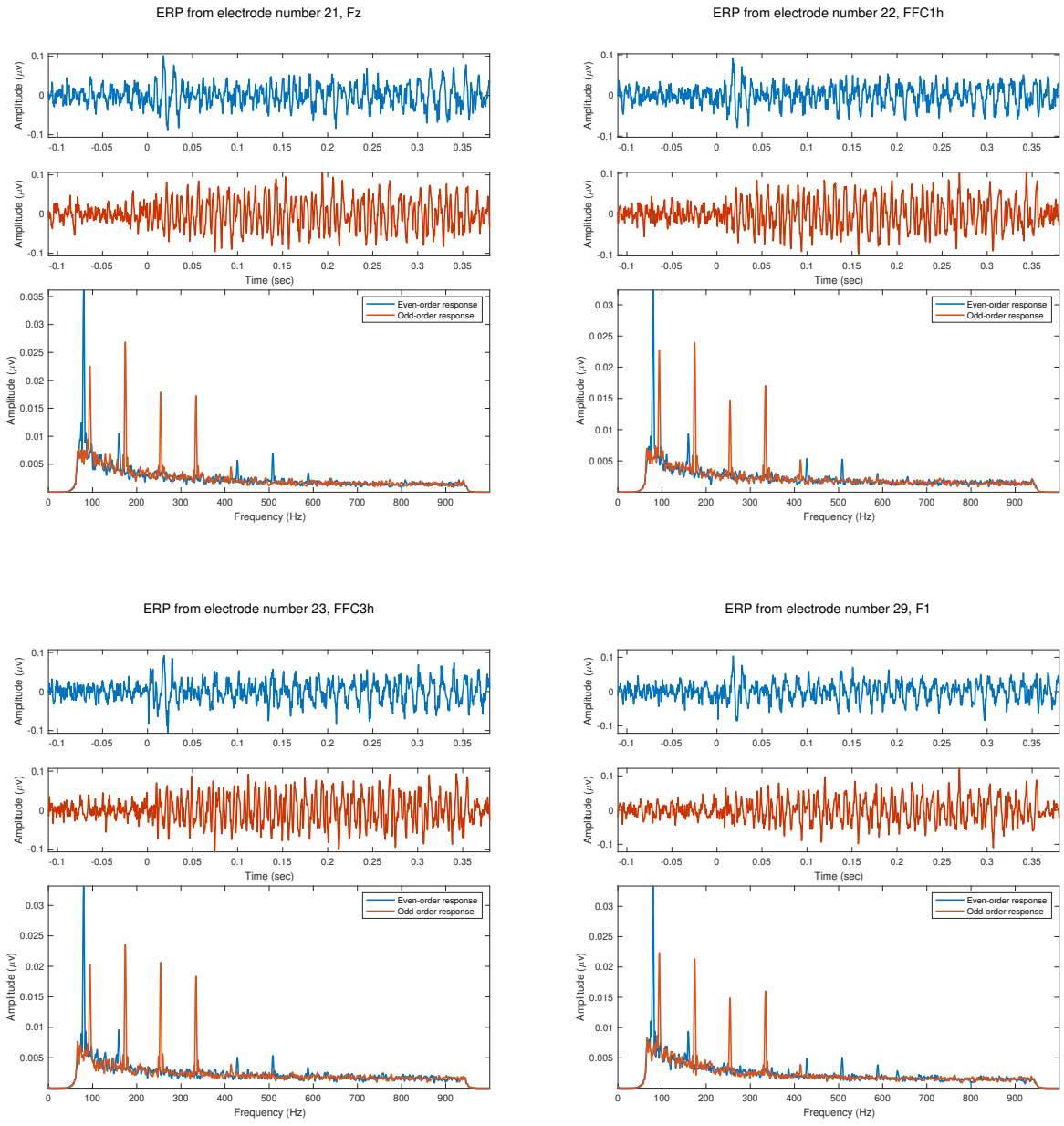


Figure 3.6: Example average scalp-space FFRs for low-f0, small shift stimulus. Time-domain waveforms are averaged in the time domain; spectra are averaged spectra.

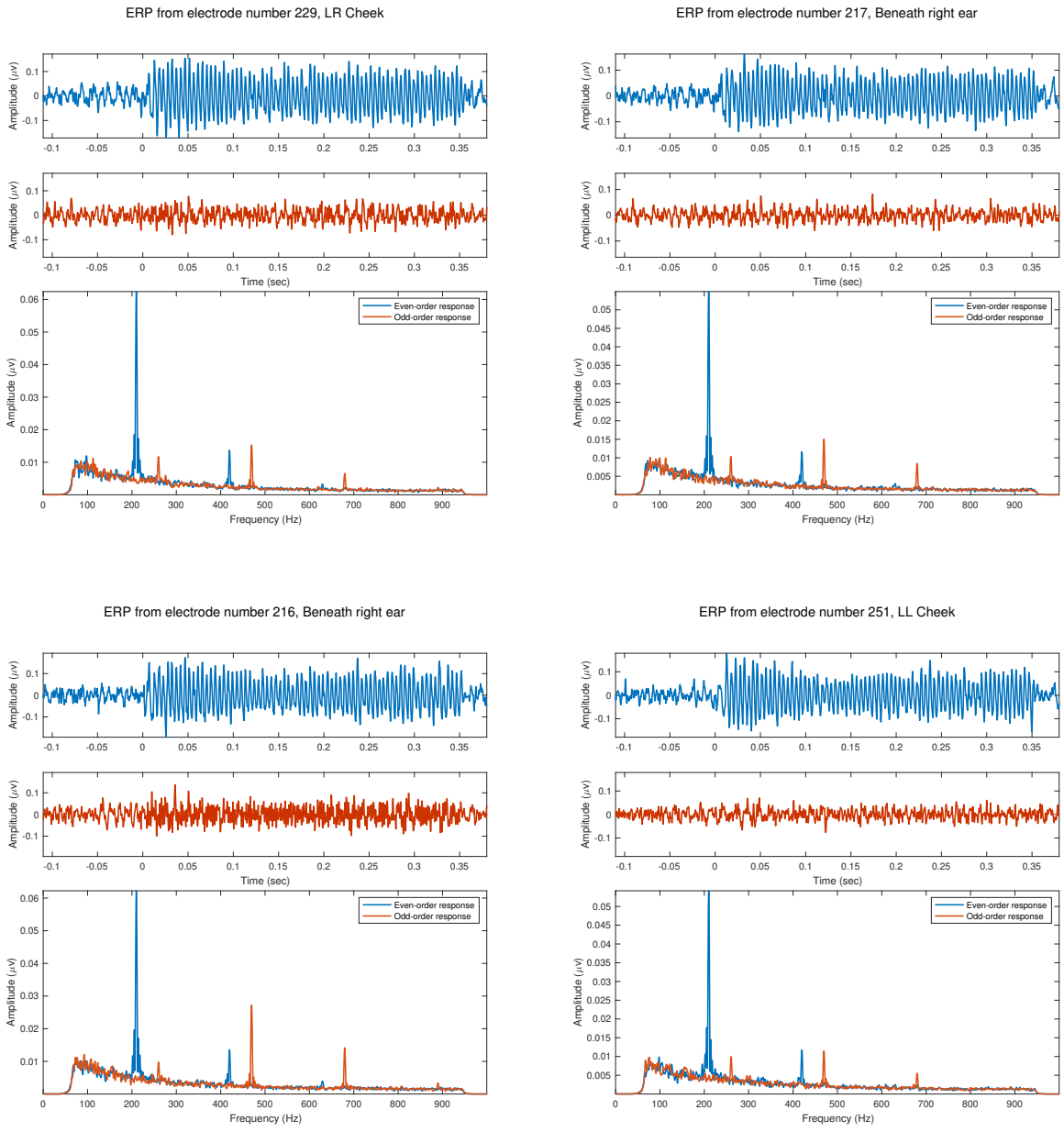


Figure 3.7: Example average scalp-space FFRs for high-f0, large shift stimulus. Time-domain waveforms are averaged in the time domain; spectra are averaged spectra.

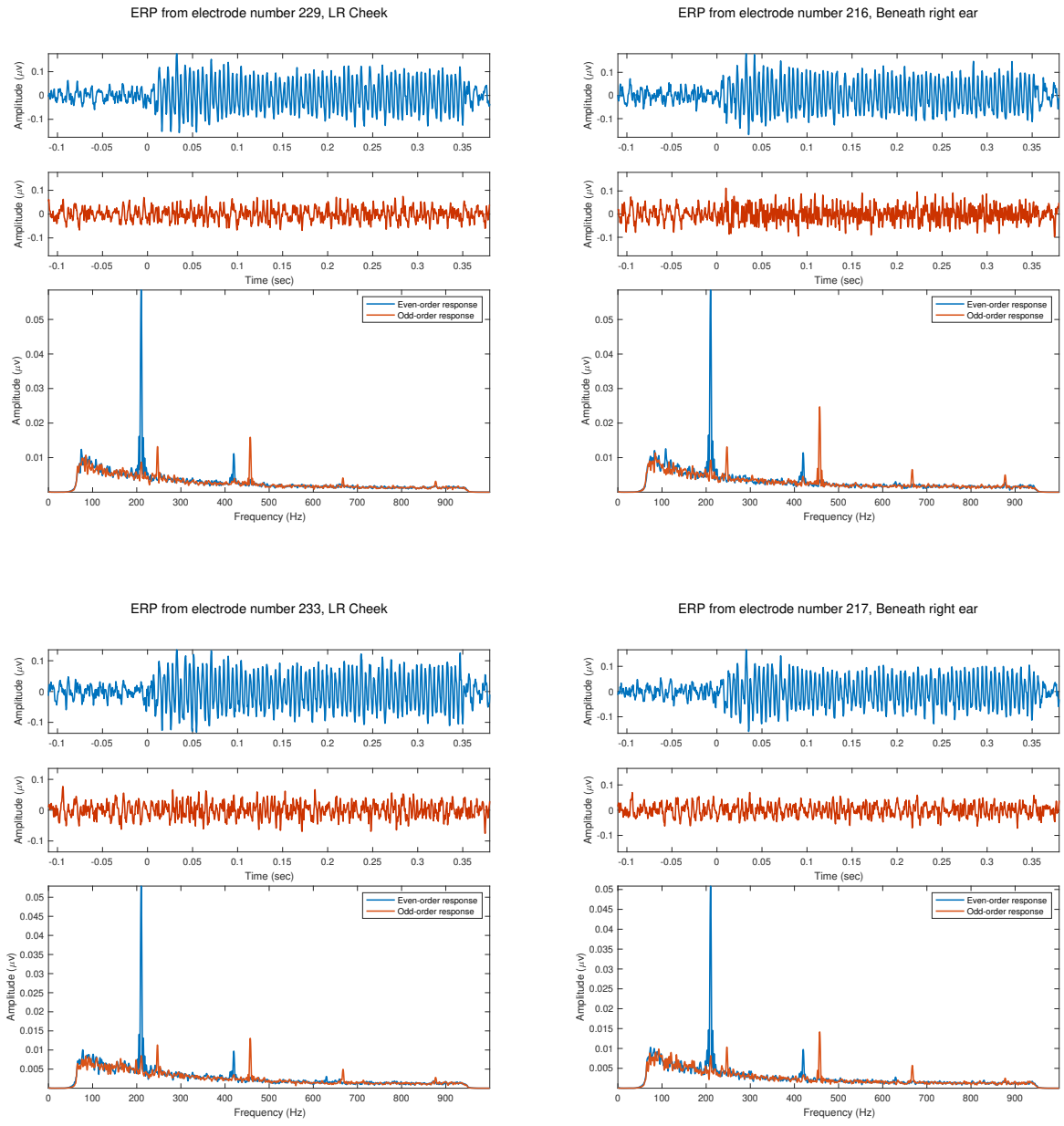
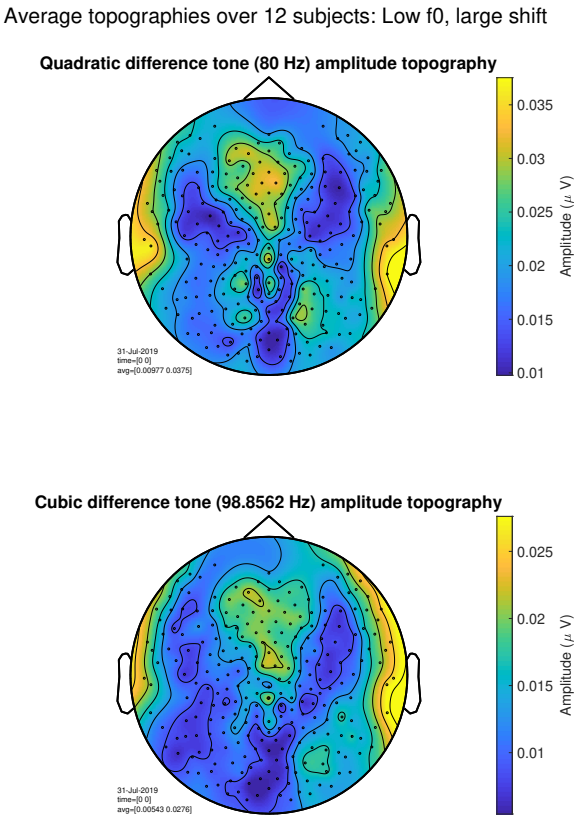


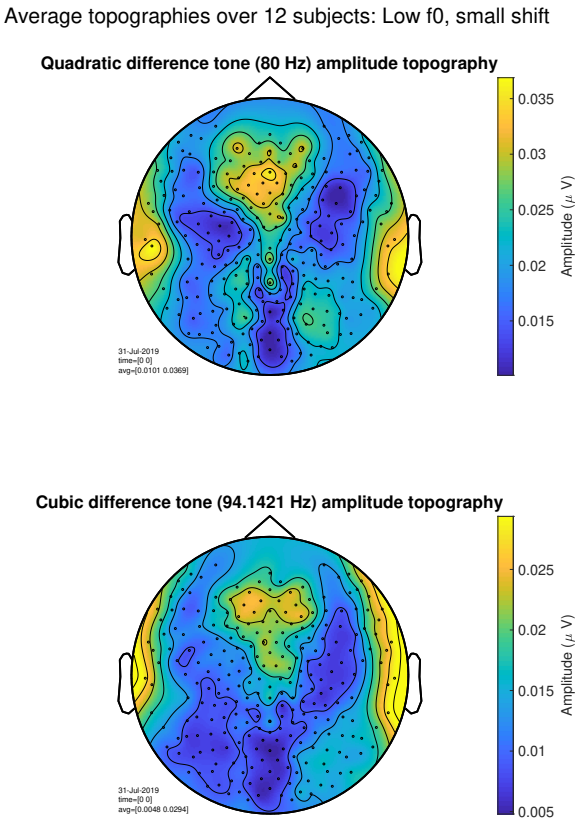
Figure 3.8: Example average scalp-space FFRs for high-f0, small shift stimulus. Time-domain waveforms are averaged in the time domain; spectra are averaged spectra.

for each of the four stimuli. These FFRs are depicted in Figures 3.5 - 3.8. A prominent QDT is present in all even-order data, and a CDT is prominent as well in the odd-order data for the low-f0 stimuli, and less prominent but still present in the high-f0 condition. There are also frequency peaks in the even-order portion that indicate non-envelope-related activity. The two or three (depending on the stimulus and electrode) higher-frequency but lower-amplitude peaks correspond to summation frequencies with respect to the three stimulus primaries, in both low-f0 cases.

The electrodes selected for display based on QDT SNR exhibit a noticeable difference between the low- and high-f0 stimuli. Electrodes around the mastoid have good SNRs for the high-f0 stimuli, while the low-f0 stimuli are more variable, but cluster more on the top, anterior portion of the head. To more completely visualize this pattern, topographical maps of the scalp were constructed to summarize the distribution of both prominent nonlinearities (QDT and CDT) around the head. Figure 3.9 shows these distributions for the low-f0 stimuli, and Figure 3.10 shows them for the high-f0 stimuli. The patterns in these scalp maps are consistent with the automated electrode choices by SNR. The locations around the mastoids were important for the high-f0 stimuli. This location is often used for a reference electrode in simple FFR studies because of its proximity to the first synapses of the auditory system. In this average-referenced data, it is clear also that this is a site of high-frequency FFR activity, whereas peaks at the low-f0 QDT are also distributed

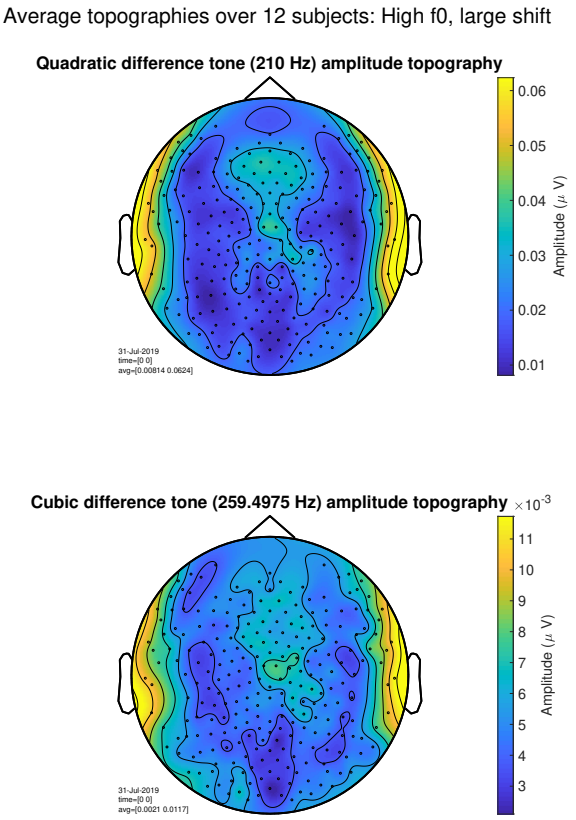


(a) Topographies for large shift condition.

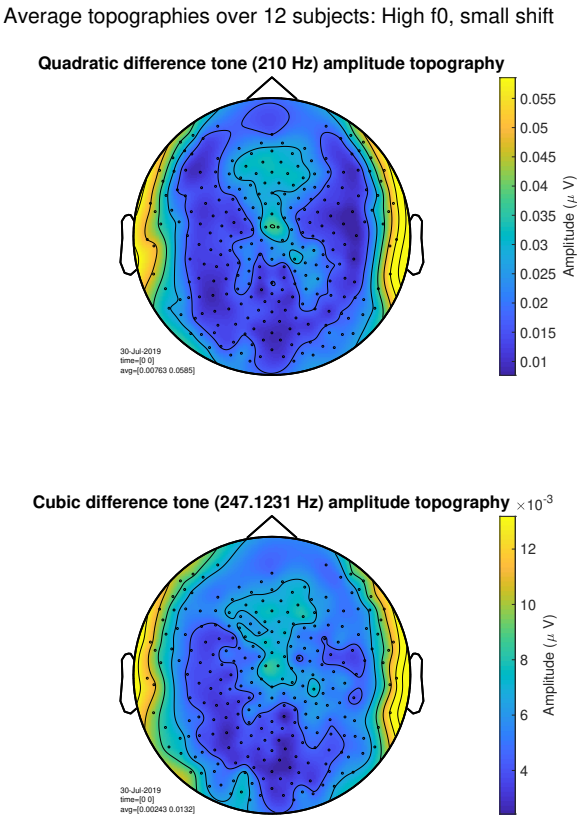


(b) Topographies for small shift condition.

Figure 3.9: Topographies of auditory nonlinearities averaged over subjects for 80 Hz f0 condition.



(a) Topographies for large shift condition.



(b) Topographies for small shift condition.

Figure 3.10: Topographies of auditory nonlinearities averaged over subjects for 210 Hz f0 condition.

around the top and front of the head, perhaps indicating more spatially-dispersed sources.

3.3.3 Source analysis of FFR

Before comparing frequency peak heights across scouts, a measure of statistical significance for the existence of spectral peaks was obtained. Source spectra were averaged across subjects for all stimulus conditions, and a mean was calculated for each spectrum with a sliding window. Each spectrum was composed of 2,845 frequency bins from 0 Hz to the Nyquist of 1,250 Hz, and the sliding window was 150 bins long. Along with the mean of each window, 2.5 standard deviations were calculated in both directions. If a frequency peak was above 2.5 standard deviations from the local mean, it was considered a significant peak and possibly subjected to further analysis. Examples of this procedure for the medial geniculate body and left auditory cortex are depicted in Figures 3.11 and 3.12. Along with the predicted QDTs, it is apparent that the three summation tones with respect to the stimulus primaries are significant for the low-f0 condition from the subcortical sources, namely the MGB in this case. These frequencies are not related to the envelope frequency, and are different depending on the amount of shift in the stimulus frequencies, whereas the QDT and its first harmonic remain the same regardless of the shift amount.

Additionally, there is a significant cortical contribution at both the low-f0 and high-f0 QDTs. The right auditory cortex (not shown) demonstrated the same significant peaks.

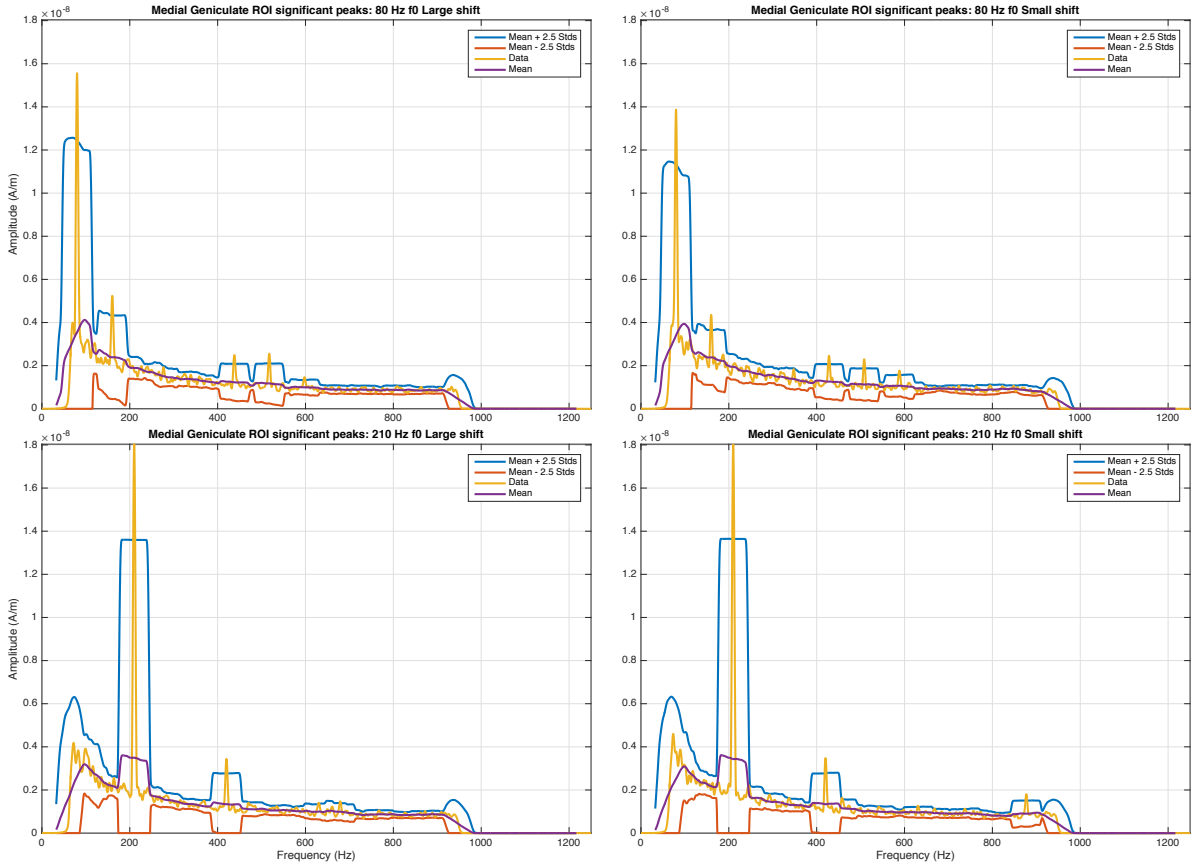


Figure 3.11: Example of peak significance testing from the MGB scout, for each of the four stimuli. Each frequency peak above 2.5 standard deviations was considered significant and could then proceed to further comparisons with other significant peaks. It is clear that the QDT, its first harmonic, and the three summation tones in the case of the lower stimuli are all significant.

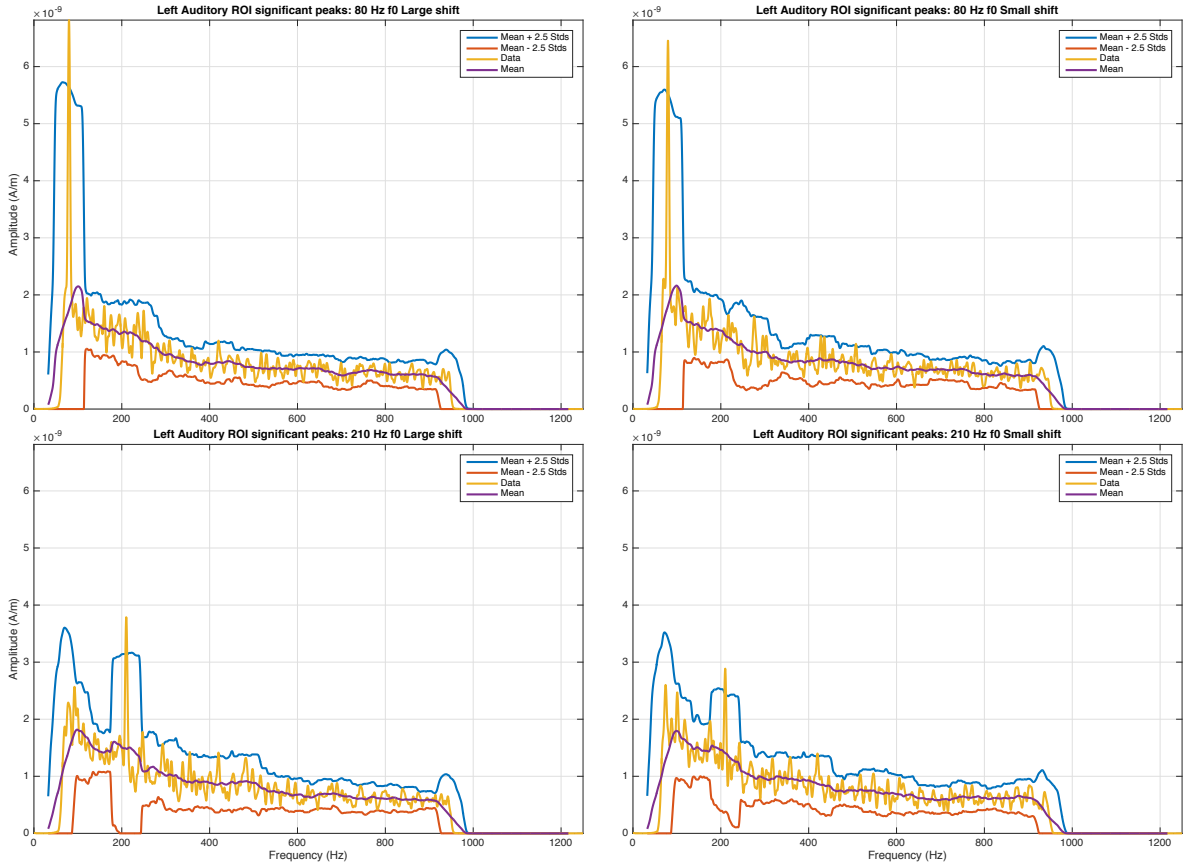


Figure 3.12: Example of peak significance testing from the left primary auditory scout, for each of the four stimuli. Each frequency peak above 2.5 standard deviations was considered significant and could then proceed to further comparisons with other significant peaks. It is clear that the QDT is significant for all stimuli.

Summaries of all even- and odd-order FFRs in source space, arranged with an emphasis on comparing cortical scouts with each other, and subcortical scouts with each other, are contained in Figures 3.13 - 3.20. There was no activity from the control cortical sources (frontal and occipital poles) for the low-f0 QDT in either shift condition, which is clear from Figure 3.13.

It is also apparent that in both shift conditions, there is a left lateralization of the QDT. Distributions of peak heights across subjects all failed the Kolmogorov-Smirnoff test of normality, hence Wilcoxon signed-rank tests for dependent means were used for all comparisons. A comparison between the left and right auditory cortex peaks (Figure 3.13) at the QDT showed a significant effect of lateralization for the large shift ($p < 0.05$, $Z = 2.20$), and a strong trend for the small shift condition ($p = 0.08$, $Z = 1.73$).

The limits of the y-axis are the same across all even-order plots for easy comparison. Figure 3.14 shows a complex spectrum from the subcortical scouts with the same features as the high-SNR scalp-space FFRs, and around twice the energy at the QDT relative to the auditory cortex sources. The QDT and its first harmonic are apparent for all four stimuli, and the non-envelope-related summation tones are also apparent for the low-f0 stimuli. MGB amplitude is the greatest of the subcortical sources, with IC in the middle and CN being the lowest. This pattern is consistent across all subcortical source analysis.

For the high-f0 stimuli, there also appears to be a trend toward left lateralization for

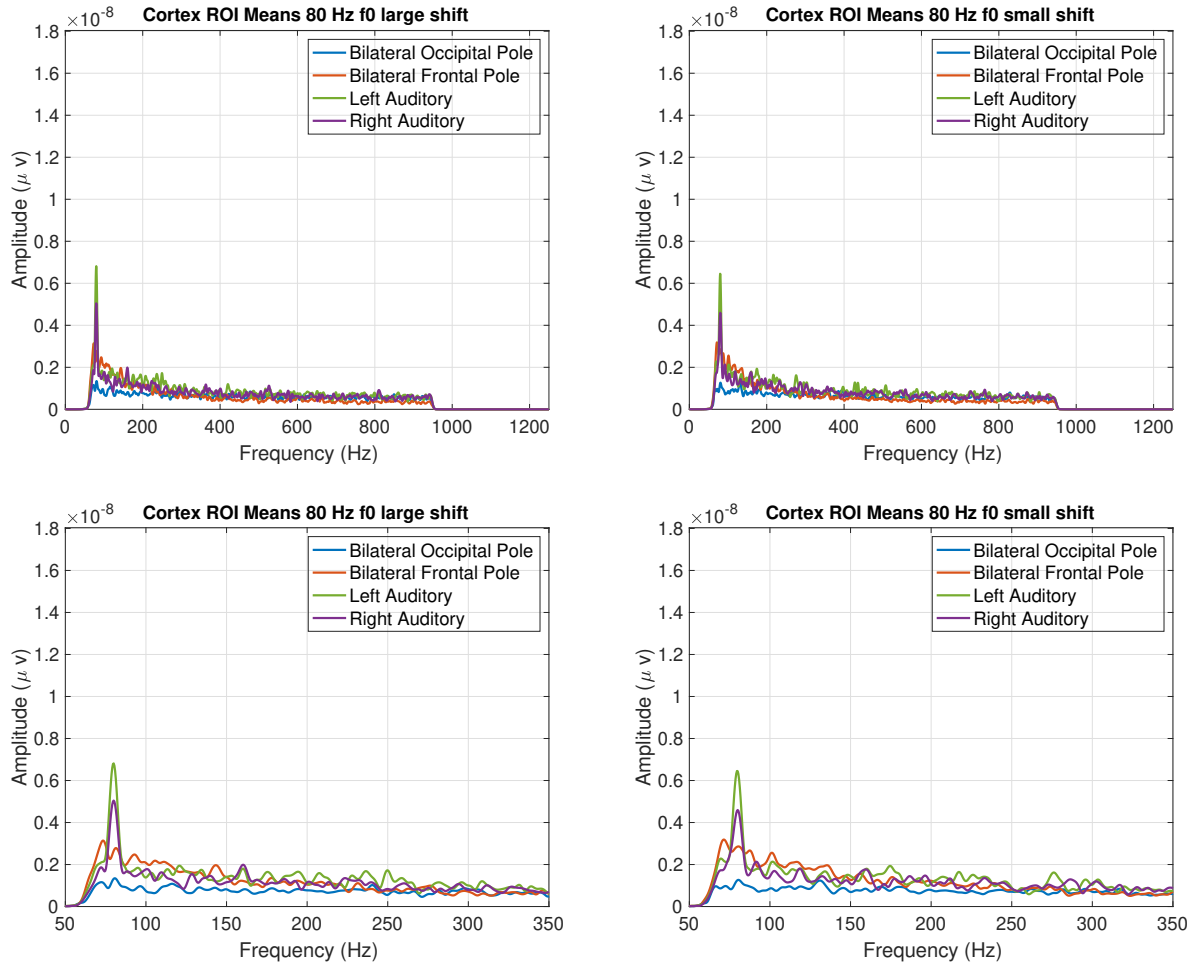


Figure 3.13: Cortical source spectra for even-order FFRs in the low-f0 condition. Lower plots are upper plots zoomed in to see detail.

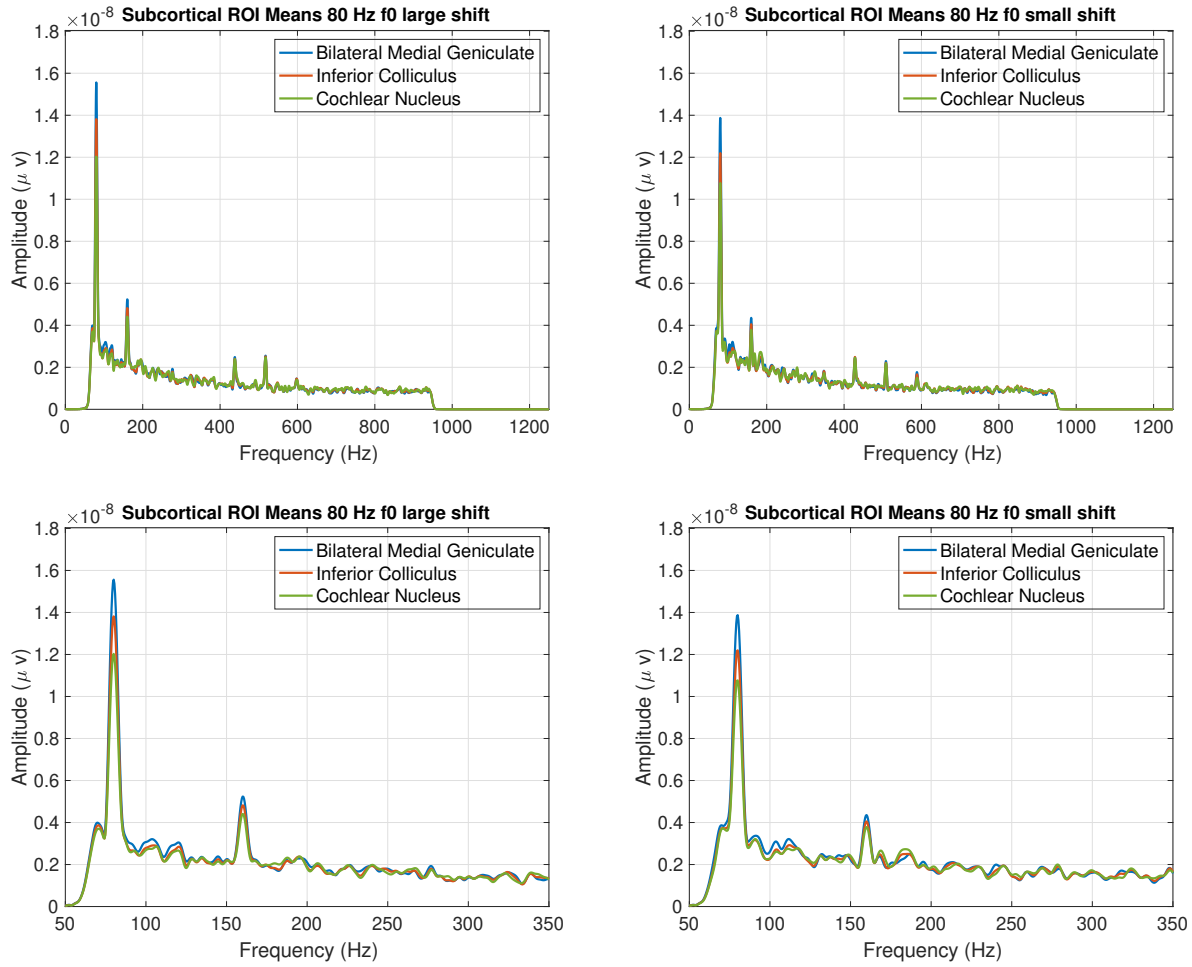


Figure 3.14: Subcortical source spectra for even-order FFRs in the low-f0 condition. Lower plots are upper plots zoomed in to see detail.

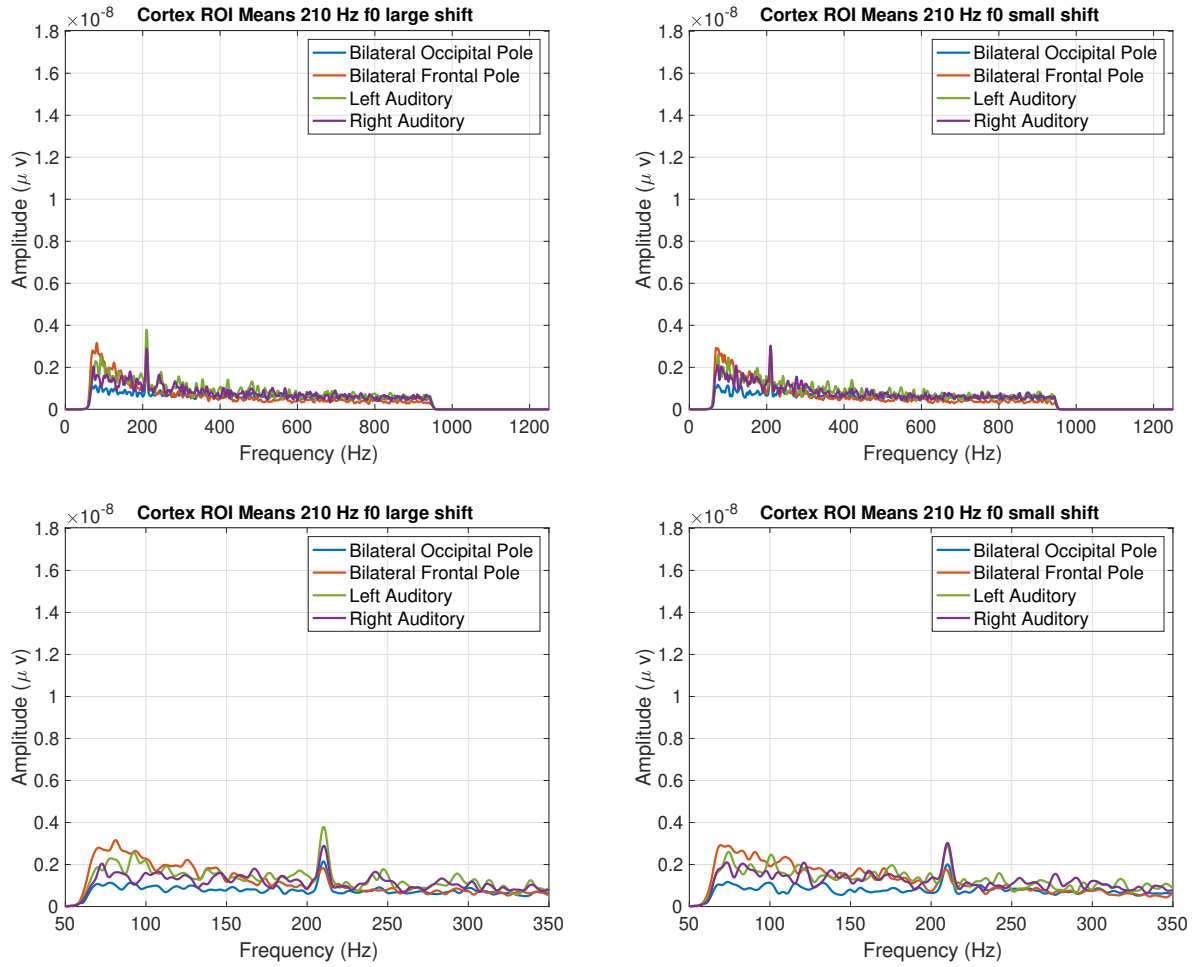


Figure 3.15: Cortical source spectra for even-order FFRs in the high-f0 condition. Lower plots are upper plots zoomed in to see detail.

cortical sources (Figure 3.15), though this was not significant for either shift condition ($p = 0.18$, $Z = 1.33$ for large shift; $p = 0.58$, $Z = -0.55$ for small shift). There are also significant peaks at the QDT from the occipital and frontal control cortical regions, which was not expected. However the auditory cortex peaks are significantly greater than the control region peaks for both shift conditions ($p < 0.05$, $Z = 2.20$ for large shift; $p < 0.005$, $Z = 2.90$ for small shift).

The same pattern of relative amplitudes in the subcortical scouts is evident for the high-f0 stimuli in Figure 3.16. Whereas the amplitude ratio of cortical and subcortical FFR sources for the low-f0 conditions was roughly $\frac{1}{2}$, the ratio here is closer to $\frac{1}{5}$, indicating a much weaker cortical contribution to the FFR for higher frequencies. This is intuitive, but not necessarily obvious. While it is true that most neurons in neocortex do not have fast enough refractory periods to fire an action potential once per cycle at e.g. 210 Hz, it is important to remember that they may still phase lock at that frequency. High phase-locking values require only consistent phase. Thus if many neurons of a population are phase-locking to an input frequency and each is firing at some subset of all input cycles, a population-level response such as the FFR would still show power at the input frequency. This is sometimes called volley theory. In this case, the input frequency is a nonlinearity generated at lower levels of the auditory system rather than a stimulus frequency, namely the QDT, but the same principles apply.

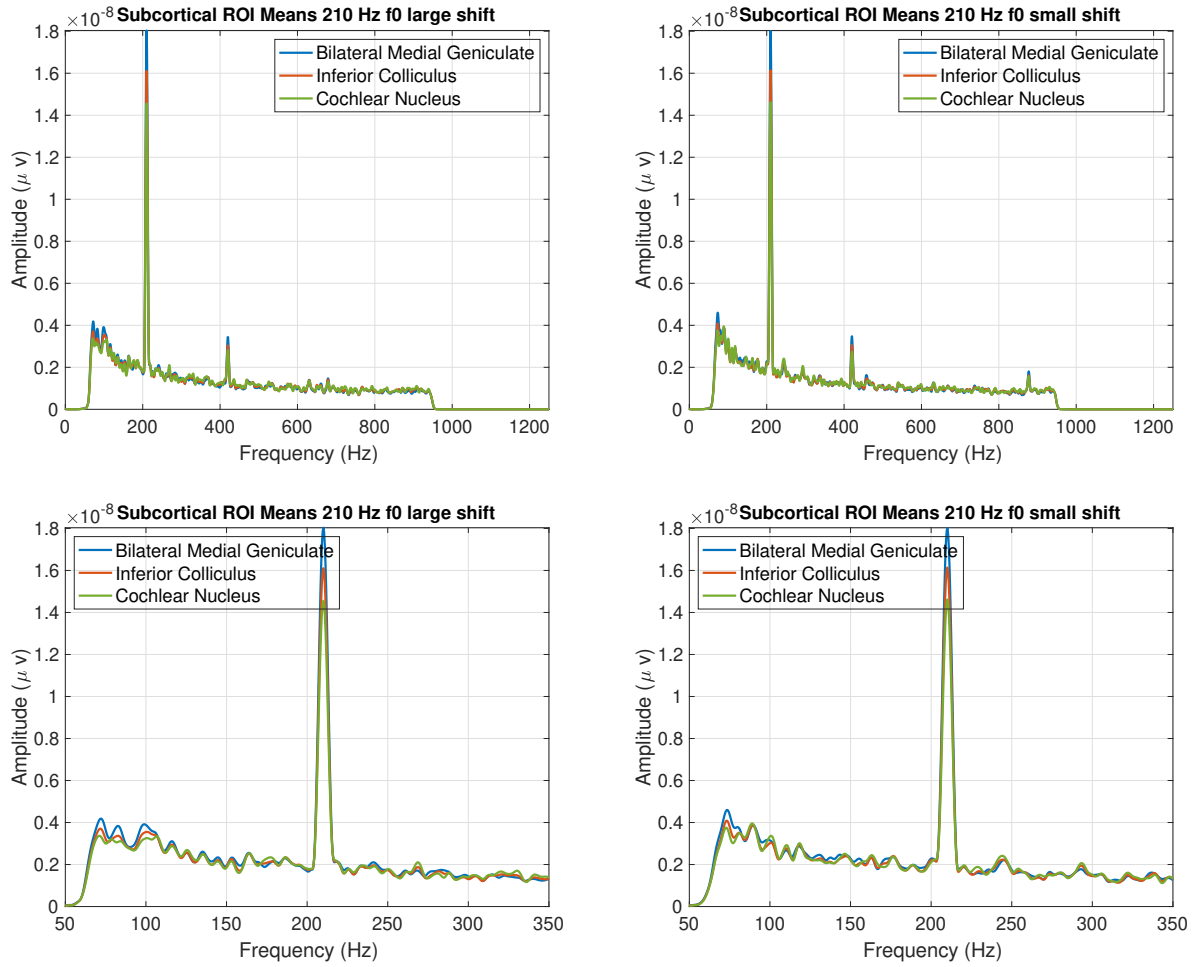


Figure 3.16: Subcortical source spectra for even-order FFRs in the high-f0 condition. Lower plots are upper plots zoomed in to see detail.

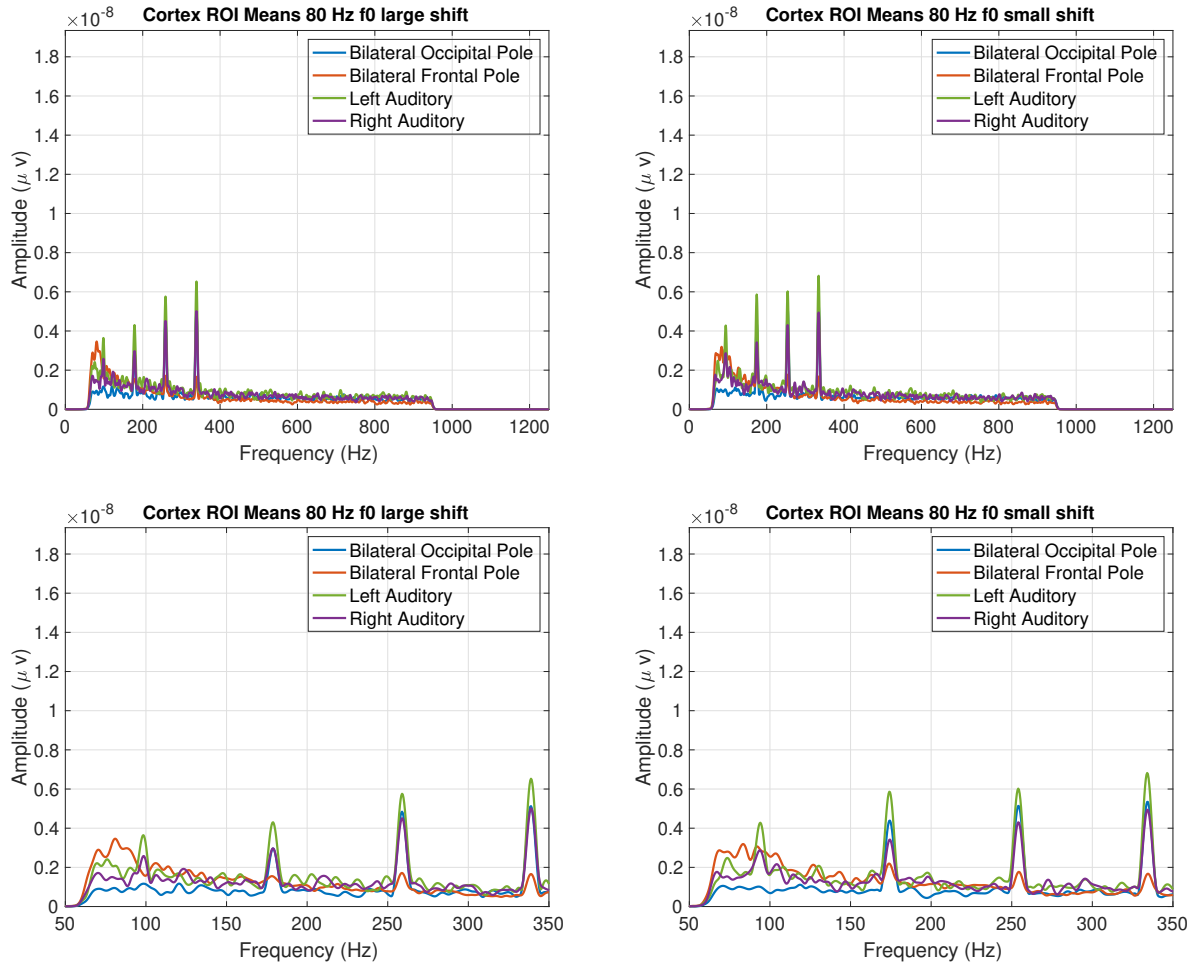


Figure 3.17: Cortical source spectra for odd-order FFRs in the low-f0 condition. Lower plots are upper plots zoomed in to see detail.

As mentioned above, this EEG data contains both trigger and stimulus artifact. Source analysis was done before the artifact removal with PCA, however the trigger artifact was avoided in the analysis pipeline by only analyzing the period after the artifact was complete. Stimulus artifact remains however, despite mu-metal shielding around the transducer. Stimulus artifact is present in the odd-order portion of the FFR to two polarity conditions, which is the subtraction of the two response averages. The odd-order source analysis provides some insight into the FFR, while also confirming stimulus artifact. This is evident in Figure 3.17, which shows the cortical source responses to the low-f0 stimuli. The chief frequency peaks in these responses are the stimulus primaries, and since they are a mixture of stimulus artifact and neural responses to the primaries, it is difficult to judge the relative contributions. The prominent responses from the occipital pole to the stimulus primaries is a strong indicator of the stimulus artifact. However the CDT is also present, and with no contribution from either control cortical source and only auditory cortex sources. Since the CDT is not a stimulus frequency and must have been generated by the brain and/or periphery, this pattern is in general an indicator that 1) the CDT is transmitted up to and is present in auditory cortex for a contribution to the FFR, and 2) stimulus artifact is contaminating this data.

The odd-order portion of the subcortical sources for the low-f0 stimuli are qualitatively similar to the even-order portion (Figure 3.18), with the MGB having the

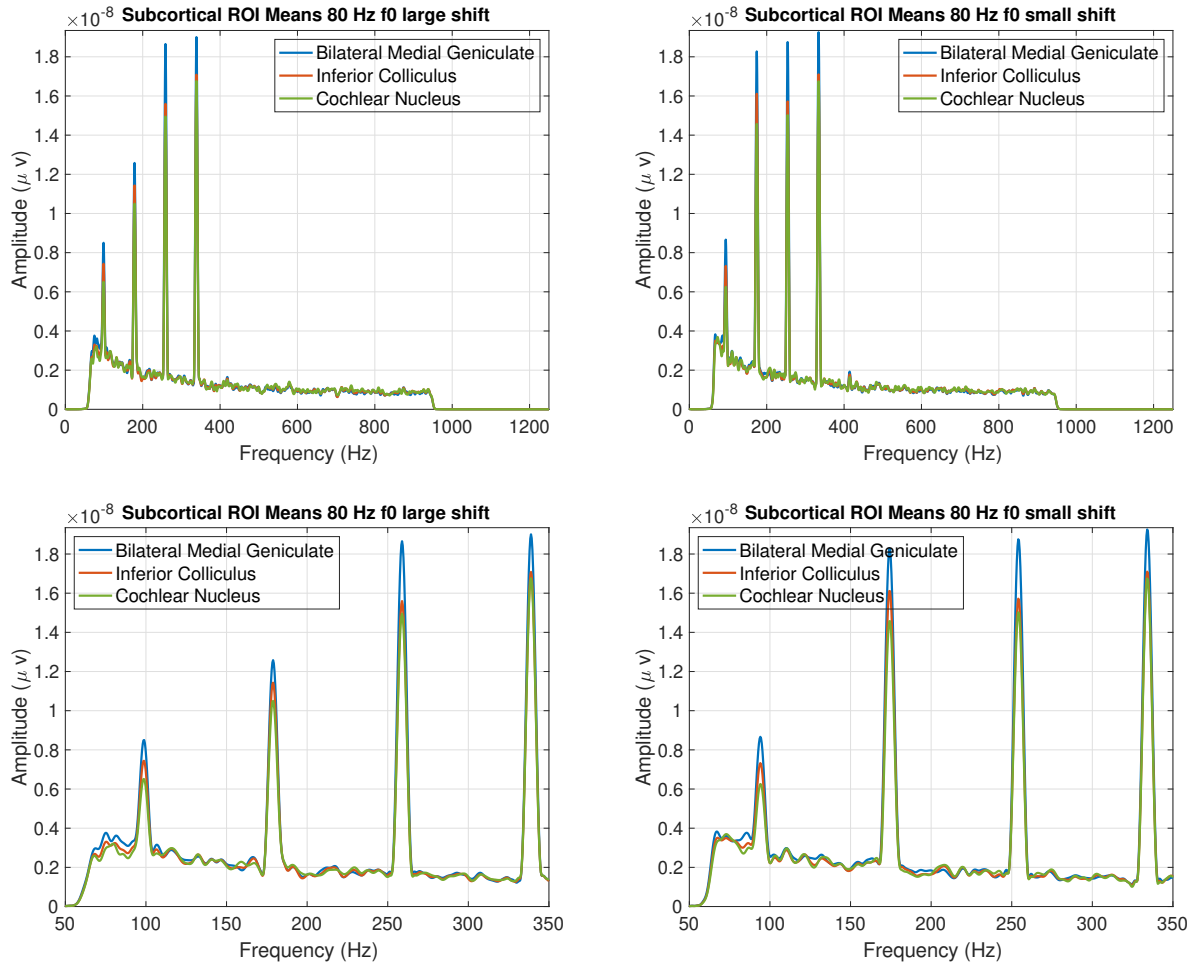


Figure 3.18: Subcortical source spectra for odd-order FFRs in the low-f0 condition. Lower plots are upper plots zoomed in to see detail.

highest amplitude, followed by the IC and then the CN sources. It is not readily possible to ascertain how much stimulus artifact is present in these responses because there were no subcortical control scouts, however it must be assumed that some exists here. A prominent CDT is present here which was expected and is by definition not artifactual, however its amplitudes are much less in general than the stimulus frequencies. Responses to the stimulus primaries are of course expected, but to be so much higher in amplitude than the CDT despite being higher in frequency indicates an artifactual component to these responses. Cortical sources of the odd-order portion for the high-f0 stimuli are shown in Figure 3.19. These responses pattern the same way as the low-f0 analog, including the activation from control sources indicating artifact. However an important difference is that there is no CDT present. It is clear that the brain and/or periphery are generating the CDT as it is visible in the subcortical sources for these stimuli (Figure 3.20), however it seems that the frequency of this nonlinearity was too high to survive up the entire auditory pathway. Even the CDT for the low-f0 condition was weak (Figure 3.18).

3.4 Discussion and concluding remarks

As more is discovered about the frequency following response, it can more readily be utilized in clinical and diagnostic settings. This includes knowing its neural and potentially

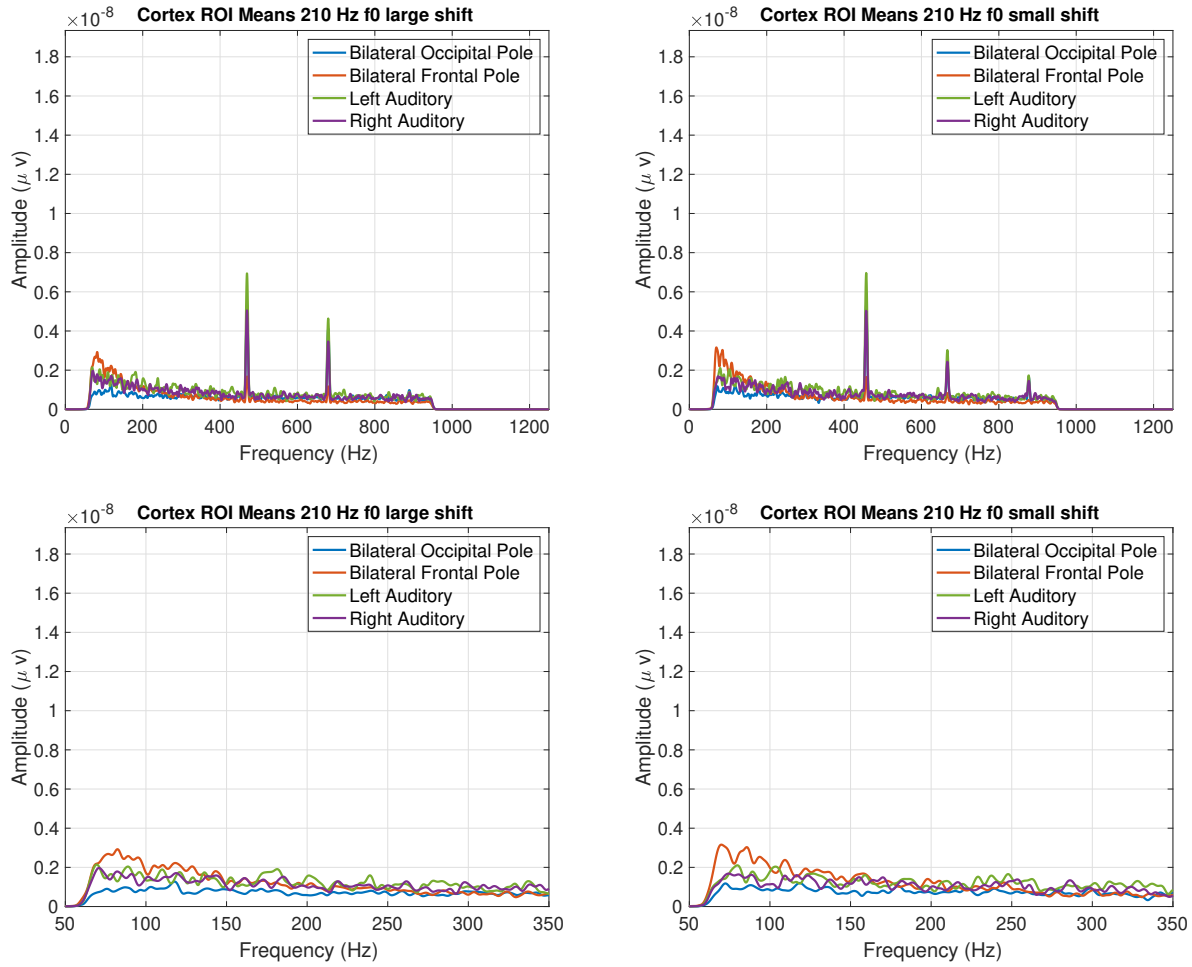


Figure 3.19: Cortical source spectra for odd-order FFRs in the high-f0 condition. Lower plots are upper plots zoomed in to see detail.

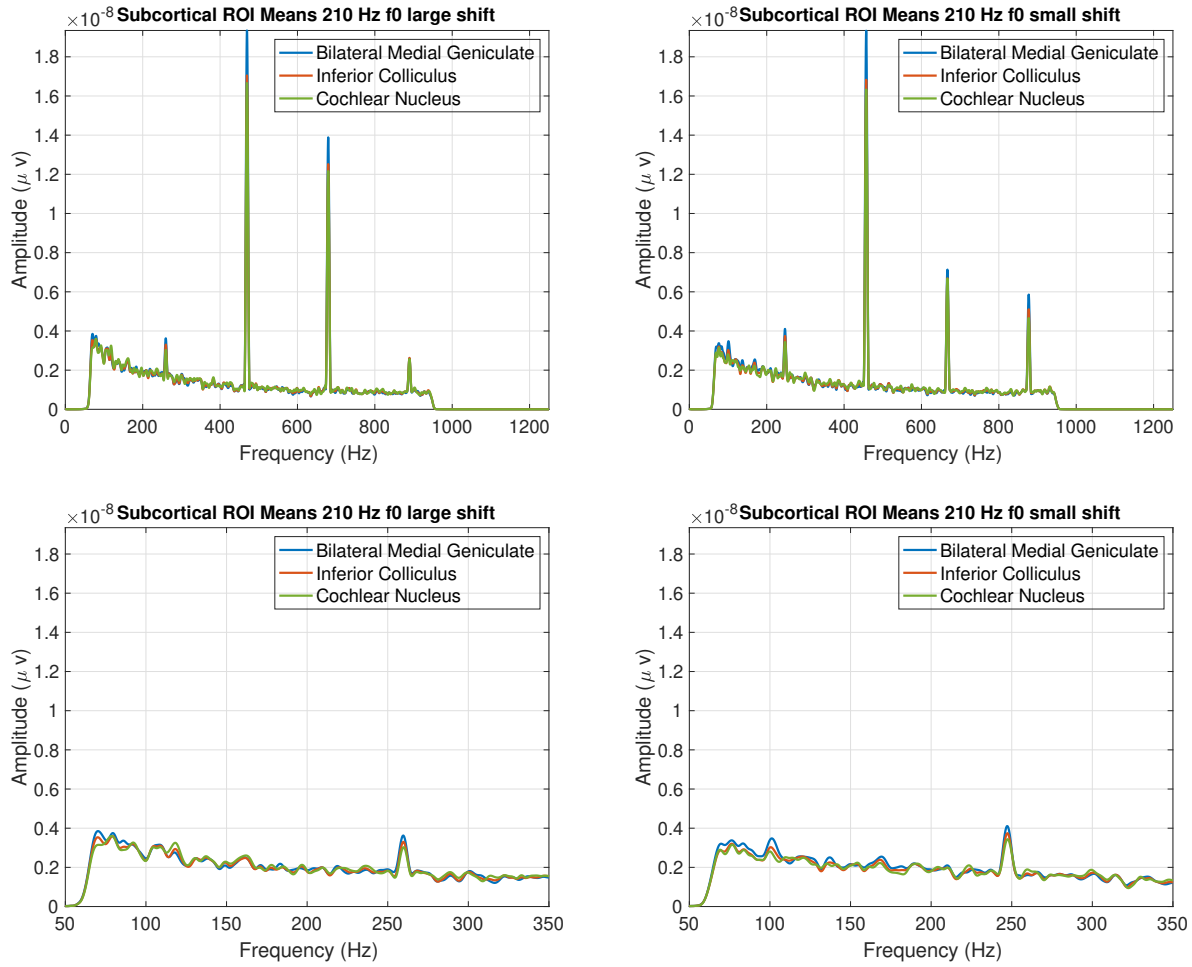


Figure 3.20: Subcortical source spectra for odd-order FFRs in the high-f0 condition. Lower plots are upper plots zoomed in to see detail.

peripheral sources. One reason for this is that it would be undesirable, all other things being equal, for frequencies in the FFR to cancel themselves out at the level of the scalp because of destructive interference from multiple sources. This is an issue that was addressed both empirically and from a modeling perspective by Tichko and Skoe (2017), in which the authors found a nonmonotonic and surprisingly richly-structured curve of pure tone FFR response amplitude as a function of frequency. They interpret this as being at least in part due to phase cancellation from multiple sources. The data from the present study suggest that two prominent generators of the FFR are the auditory thalamus and primary auditory cortex. If the time delay between these two synapses is equal to one half of the reciprocal of a prominent FFR frequency (which may or may not be a stimulus frequency itself), then destructive interference from these two sources may be expected. This is because the response at this frequency would be at opposing phases in the two structures at a given time.

The present source analysis fits well with the data of Tichko and Skoe (2017) if their own approximations of delay times are utilized. This source analysis shows a prominent response from primary auditory cortex (AC), as well as from the auditory thalamus (MGB) and the inferior colliculus (IC). The approximate delay times for MGB to AC, and IC to AC, are 5 milliseconds and 7 milliseconds respectively. These correspond to maximally destructive interference for the frequencies 100 Hz and ≈ 71.43 Hz respectively. And indeed

there are troughs in their data curves at approximately these frequencies, with a peak in between them.

From the standpoint of pitch perception, the present work is an extension of Gockel et al. (2011) and Greenberg et al. (1987) before them. When one utilizes harmonic stimuli such as a speech syllable, one certainly observes the fundamental frequency prominently in the FFR. For harmonic stimuli, it is also appropriate to refer to this as the envelope frequency, the difference frequency, and the pitch frequency. This is all also true for a harmonic missing fundamental-type stimulus. It is well known that the pitch frequency is still at the missing fundamental in this case. It is also the case that the FFR prominently reconstructs this frequency, thus it was thought for a period of time that the FFR contained a direct correlate of pitch. Using pitch-shifted stimuli, it was demonstrated here that the FFR in fact does not in general contain pitch frequencies, but instead is comprised of both even- and odd-order nonlinearities explained as combination tones with respect to the stimulus primaries and other FFR frequencies. The pitch frequency is instead predicted directly from the stimulus with Equation 2.7, or with the autocorrelation of the stimulus itself or of the odd-order portion of the FFR (Gockel et al., 2011).

This study also has clarified the notion of "envelope" as it relates to the FFR. The FFR, or at least the even-order portion of it, is often referred to as the "envelope-following response". The present stimuli were chosen because, for a given QDT (or original

fundamental frequency before shifting), different shift conditions have identical Hilbert amplitude envelopes. Thus if the even-order portion of the FFR is an envelope-following response, these should be identical across shift conditions for a given f0. But this data shows that, while the QDT (envelope frequency) and its first harmonic are identically present in both shift conditions, there are also even-order frequency components in the FFRs that are not related to the envelope frequency, and thus are indeed different across shift conditions. These even-order components can be explained as nonlinear summation responses with respect to the stimulus primary frequencies. In FFRs with better SNR, there may also be similar nonlinearities predicted as combination tones with respect to FFR frequencies generated by the brain that are not present in the stimulus.

Observing Figures 3.9 and 3.10, it is clear that there is not an appreciable difference between the QDT and CDT in terms of their scalp-space topography. This does not necessarily indicate that they have the same generators, however. As reviewed above, lower-level physiological studies of the periphery indicate strongly that odd-order nonlinearity is generated largely in the cochlea and is then transmitted neurally up the auditory system. When the CDT frequency is subsequently located in the post-synaptic potentials of neurons at various stages, it can then be found in the FFR.

A comparison between low- and high-f0 stimuli, on the other hand, shows both scalp-space and source-space distinctions in the FFR. The topography comparisons suggest

most activity for the high-f0 conditions is coming from the lower brainstem, while the low-f0 responses suggest more thalamic, cortical, or otherwise higher-level contribution. And indeed the source analysis bears this out; Figure 3.13 demonstrates a cortical contribution to the FFR at the QDT for the low-f0 stimuli. Figure 3.15 shows a small but significant contribution from cortex for the high-f0 responses as well, but with the ratio of subcortical-to-cortical contribution much greater for the responses to those stimuli.

While similar studies have been undertaken in recent years and months, this is the first study to utilize high-density EEG and structural MRI to conduct a source analysis on the FFR to multiple complex tones. The results provide more converging evidence of a cortical contribution to the FFR under most electrode montage regimes, as was also prominently concluded by Coffey et al. (2016). However there are multiple aspects of this study that need to be replicated. Firstly, confirmation of the various auditory nonlinearities expected as combination tones with respect both to stimulus frequencies and other FFR frequencies is necessary for a more comprehensive understanding of the FFR's frequency content. This can clearly be much more readily achieved with lower impedances than those used out of necessity in the present study. A low-impedance, high-density EEG system would be prohibitively slow to prepare. However, gel-based, low-impedance EEG systems exist with 128 electrodes. A replication attempt of the present source analysis results with such a system would be in order. There is evidence that the inclusion of

inferiorly-located electrodes such as those on the cheeks, below the ears, and on the neck in EGI's 256-electrode net contributes strongly to the ability to localize scalp-space signals to deeper structures such as the brainstem (Song et al., 2015). It is an open question what a source analysis excluding these electrode placements would show with regard to the FFR.

The chief limitations of the present study are clear: High impedances, the presence of trigger artifact, and the presence of stimulus artifact. While the impedance problem is difficult to solve if one wants a dense sampling of the scalp space, various artifacts can be vigorously controlled for in future studies. The presence of the trigger artifact unfortunately meant that a crucial part of the analysis in Coffey et al. (2016) could not be replicated, namely the delay-based identification of successive auditory structures. The onset responses from most of the auditory system are over well before 60 milliseconds, which was the length of the trigger artifact in each trial here. Fortunately frequency analysis could still be done on the remainder of the FFR, but more detailed study of the combined onset responses from successive structures is highly desirable for the future.

References

Abel, C., Wittekindt, A., & Kössl, M. (2009). Contralateral acoustic stimulation modulates low-frequency biasing of DPOAE: Efferent influence on cochlear amplifier operating state? *Journal of Neurophysiology*, *101*(5), 2362–71. doi:10.1152/jn.00026.2009

Ahlfors, S. P., Han, J., Belliveau, J. W., & Hämäläinen, M. S. (2010). Sensitivity of MEG and EEG to source orientation. *Brain Topography*, *23*(3), 227–232. doi:10.1007/s10548-010-0154-x

Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology & neuro-otology*, *11*(4), 213–32. doi:10.1159/000092589

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1-2), 35–47. doi:10.1016/j.heares.2008.08.004

Althen, H., Wittekindt, A., Gaese, B., Kössl, M., & Abel, C. (2012). Effect of contralateral pure tone stimulation on distortion emissions suggests a frequency-specific

functioning of the efferent cochlear control. *Journal of Neurophysiology*, *107*(7), 1962–9. doi:10.1152/jn.00418.2011

Arnold, S., & Burkard, R. (1998). The auditory evoked potential difference tone and cubic difference tone measured from the inferior colliculus of the chinchilla. *Journal of the Acoustical Society of America*, *104*(3 Pt 1), 1565–73.

Arnold, S., & Burkard, R. (2000). Studies of interaural attenuation to investigate the validity of a dichotic difference tone response recorded from the inferior colliculus in the chinchilla. *Journal of the Acoustical Society of America*, *107*(3), 1541. doi:10.1121/1.428439

Bhagat, S. P., & Champlin, C. A. (2004). Evaluation of distortion products produced by the human auditory system. *Hearing Research*, *193*(1-2), 51–67. doi:10.1016/j.heares.2004.04.005

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research*, *323*(May), 68–80. doi:10.1016/j.heares.2015.01.011

Buunen, T. J., & Rhode, W. S. (1978). Responses of fibers in the cat's auditory nerve to the cubic difference tone. *Journal of the Acoustical Society of America*, *64*(3), 772–781. doi:10.1121/1.382042

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, *76*(3).

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, *47*(2), 236–46. doi:10.1111/j.1469-8986.2009.00928.x

Chertoff, M. E., & Hecox, K. E. (1990). Auditory nonlinearities measured with auditory-evoked potentials. *Journal of the Acoustical Society of America*, *87*(3), 1248–54.

Chertoff, M. E., Hecox, K. E., & Goldstein, R. (1992). Auditory distortion products measured with averaged auditory evoked potentials. *Journal of Speech and Hearing Research*, *35*(1), 157–66.

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications*, *7*, 1–11. doi:10.1038/ncomms11070

Cooper, N. P. (1998). Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea. *The Journal of physiology*, *509*(1), 277–88.

de Boer, E. (1956). On the 'residue' in hearing. PhD Thesis. *Academisch Proefschrift, Universiteit van Amsterdam, Amsterdam.*

de Cheveigné, A. (2005). Pitch Perception Models. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). doi:10.1007/0-387-28958-5_6

de Cheveigné, A., & Pressnitzer, D. (2006). The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. *Journal of the Acoustical Society of America*, *119*(6).

Deeter, R., Abel, R., Calandruccio, L., & Dhar, S. (2009). Contralateral acoustic stimulation alters the magnitude and phase of distortion product otoacoustic emissions. *Journal of the Acoustical Society of America*, *126*(5), 2413–24. doi:10.1121/1.3224716

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. doi:10.1016/j.neuroimage.2006.01.021

Dolphin, W. F., & Mountain, D. C. (1992). The envelope following response: Scalp potentials elicited in the mongolian gerbil using sinusoidally AM acoustic signals. *Hearing Research*, *58*, 70–78. doi:10.1016/0378-5955(92)90010-K

Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., & Magnasco, M. O. (2000). Essential nonlinearities in hearing. *Physical Review Letters*, *84*(22), 5232–5.

Elsisy, H., & Krishnan, A. (2008). Comparison of the acoustic and neural distortion product at 2f1-f2 in normal-hearing adults. *International Journal of Audiology*, *47*(7), 431–8. doi:10.1080/14992020801987396

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–81. doi:10.1016/j.neuroimage.2012.01.021

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., ... Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–55.

Fletcher, H. (1924). The physical criterion for determining the pitch of a musical tone. *Physical Review*, *23*(3), 427–437. doi:10.1103/PhysRev.23.427

Fletcher, H. (1929). *Speech and hearing*. New York: D. Van Nostrand Company, Inc.

Gardi, J., Merzenich, M., & McKean, C. (1979). Origins of the scalp recorded frequency-following response in the cat. *Audiology: Official organ of the International Society of Audiology*, *18*(5), 358–81.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology*, *12*(6), 767–82. doi:10.1007/s10162-011-0284-1

Goldenholz, D. M., Ahlfors, S. P., Hämäläinen, M. S., Sharon, D., Ishitobi, M., Vaina, L. M., & Stufflebeam, S. M. (2008). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Human Brain Mapping*, *30*(4), 1077–1086. doi:10.1002/hbm.20571

Goldstein, J. L. (1970). Aural combination tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 230–245). A.W Sijthoff, Leiden.

Goldstein, J. L., & Kiang, N. (1968). Neural Correlates of the Aural Combination Tone 2f1-f2. *Proceedings of the IEEE*, *56*(6), 981–992. doi:10.1109/PROC.1968.6449

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–60. doi:10.1016/j.neuroimage.2013.10.027

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMedical Engineering OnLine*, *9*(45), 1–20. doi:10.1186/1475-925X-8-1

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2011). Forward Field Computation with OpenMEEG. *Computational Intelligence and Neuroscience*, *2011*, 1–13. doi:10.1155/2011/923703

Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research*, *25*(2-3), 91–114.

Hudspeth, A. J., Jülicher, F., & Martin, P. (2010). A critique of the critical cochlea: Hopf–a bifurcation–is better than none. *Journal of Neurophysiology*, *104*(3), 1219–29. doi:10.1152/jn.00437.2010

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, *34*(4), 1600–1611. doi:10.1016/j.neuroimage.2006.09.024

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America*, *64*(5), 1386–1391. doi:10.1121/1.382104

Kemp, D. T. (1979). Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. *Archives of Oto-Rhino-Laryngology*, *224*(1-2), 37–45. doi:10.1007/BF00455222

Kim, D. O., Molnar, C. E., & Matthews, J. W. (1980). Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure. *Journal of the Acoustical Society of America*, *67*(5), 1704–21.

Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neuro-Otology*, *4*(2), 95–103. doi:10.1159/000013826

Kujawa, S., Fallon, M., & Bobbin, R. P. (1995). Time-varying alterations in the f2-f1 DPOAE response to continuous primary stimulation I: Response characterization and contribution of the olivocochlear efferents. *Hearing Research*, *85*, 142–154.

Kujawa, S., & Liberman, M. C. (2001). Effects of olivocochlear feedback on distortion product otoacoustic emissions in guinea pig. *Journal of the Association for Research in Otolaryngology*, *2*, 268–278. doi:10.1007/s101620010047

Laudanski, J., Coombes, S., Palmer, A. R., & Sumner, C. J. (2010). Mode-locked spike trains in responses of ventral cochlear nucleus chopper and onset neurons to periodic stimuli. *Journal of Neurophysiology*, *103*(3), 1226–37. doi:10.1152/jn.00070.2009

Lerud, K. D., Almonte, F. V., Kim, J. C., & Large, E. W. (2014). Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hearing Research*, *308*, 41–9. doi:10.1016/j.heares.2013.09.010

Licklider, J. C. R. (1956). Auditory frequency analysis. In C. Cherry (Ed.), *Information theory* (pp. 253–268). New York, NY: Academic Press.

Moore, G. A., & Moore, B. C. J. (2003). Perception of the low pitch of frequency-shifted complexes. *Journal of the Acoustical Society of America*, *113*(2), 977–985. doi:10.1121/1.1536631

Moushegian, G., Rupert, A., & Stillman, R. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, *35*(6), 665–667.

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. doi:10.1016/j.jneumeth.2010.07.015

Nuttall, A. L., & Dolan, D. F. (1993). Intermodulation distortion (F2-F1) in inner hair cell and basilar membrane responses. *Journal of the Acoustical Society of America*, *93*(4), 2061–2068.

Nuttall, A. L., Dolan, D. F., & Avinash, G. (1990). Measurements of basilar membrane tuning and distortion with laser Doppler velocimetry. In P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, & C. R. Steele (Eds.), *The mechanics and biophysics of hearing* (pp. 288–295). doi:10.1007/978-1-4757-4341-8_35

Ohm, G. S. (1843). Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Annalen der Physik*, *135*(8), 513–565. doi:10.1002/andp.18431350802

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. doi:10.1155/2011/156869

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*, 713–719.

Patterson, R. D. (1973). The effects of relative phase and the number of components on residue pitch. *Journal of the Acoustical Society of America*, *53*(6), 1565–1572.

Rickman, M., Chertoff, M., & Hecox, K. (1991). Electrophysiological evidence of nonlinear distortion products to two-tone stimuli. *Journal of the Acoustical Society of America*, *89*(6), 2818–2826.

Robles, L., Ruggero, M. A., & Rich, N. C. (1990). Two-Tone Distortion Products in the Basilar Membrane of the Chinchilla Cochlea. In P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, & C. R. Steele (Eds.), *The mechanics and biophysics of hearing* (pp. 304–313). doi:10.1007/978-1-4757-4341-8_37

Robles, L., Ruggero, M. A., & Rich, N. C. (1997). Two-Tone Distortion on the Basilar Membrane of the Chinchilla Cochlea. *Journal of Neurophysiology*, *77*, 2385–2399.

Ruggero, M. A., & Rich, N. C. (1991). Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration. *Hearing Research*, *51*(2), 215–230. doi:10.1016/0378-5955(91)90038-B

Schofield, B. R., & Cant, N. B. (1996). Projections from the ventral cochlear nucleus to the inferior colliculus and the contralateral cochlear nucleus in guinea pigs. *Hearing Research*, *102*(1-2), 1–14. doi:10.1016/S0378-5955(96)00121-9

Schouten, J. F. (1940a). The perception of pitch. *Philips Technical Review*, *5*(10), 286–294.

Schouten, J. F. (1940b). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Nederlandsche Akademie von Wetenschappen*, *43*, 991–999.

Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, *294*(1940), 1418–1424.

Schroeder, M. R. (1966). Residue Pitch : A Remaining Paradox and a Possible Explanation. *Journal of the Acoustical Society of America*, *40*(1), 79–81.

Seebeck, A. (1841). Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Annalen der Physik*, *129*(7), 417–436. doi:10.1002/andp.18411290702

Skoe, E., Burakiewicz, E., Figueiredo, M., & Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience*, *349*, 278–290. doi:10.1016/j.neuroscience.2017.02.049

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds : A Tutorial. *Ear and Hearing*, *31*(3), 1–23.

Smalt, C. J., Krishnan, A., Bidelman, G. M., Ananthakrishnan, S., & Gandour, J. T. (2012). Distortion products and their influence on representation of pitch-relevant

information in the human brainstem for unresolved harmonic complex tones. *Hearing Research*, *292*(1-2), 26–34. doi:10.1016/j.heares.2012.08.001

Smith, J. C., Marsh, J. T., Greenberg, S., & Brown, W. S. (1978). Human auditory frequency-following responses to a missing fundamental. *Science (New York, N.Y.)* *201*(4356), 639–641. doi:10.1126/science.675250

Smoorenburg, G. F. (1970). Pitch perception of two-frequency stimuli. *Journal of the Acoustical Society of America*, *48*(4), 924–42.

Sohmer, H., Pratt, H., & Kinarti, R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology*, *42*, 656–664.

Song, J., Davey, C., Poulsen, C., Luu, P., Turovets, S., Anderson, E., ... Tucker, D. (2015). EEG source localization: Sensor density and head surface coverage. *Journal of Neuroscience Methods*, *256*, 9–21. doi:10.1016/j.jneumeth.2015.08.015

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*. doi:10.1155/2011/879716

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*(5), 1061–9.

Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hearing Research*, *348*, 1–15. doi:10.1016/j.heares.2017.01.014

von Helmholtz, H. (1863). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (1st ed.). Braunschweig: Druck und Verlag von Friedrich Vieweg und Sohn.

Walliser, K. (1969). Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe. *Acustica*, *21*, 319–328.

Wever, E. G., & Bray, C. W. (1930a). Action currents in the auditory nerve in response to acoustical stimulation. *Proceedings of the National Academy of Sciences*, *16*, 344–350.

Wever, E. G., & Bray, C. W. (1930b). The nature of the acoustic response: The relation between sound frequency and frequency of impulses in the auditory nerve. *Journal of Experimental Psychology*, *8*(5), 373–387.

Wile, D., & Balaban, E. (2007). An auditory neural correlate suggests a mechanism underlying holistic pitch perception. *PLoS ONE*, *2*(4), e369. doi:10.1371/journal.pone.0000369

Wittekindt, A., Gaese, B. H., & Kössl, M. (2009). Influence of contralateral acoustic stimulation on the quadratic distortion product f2-f1 in humans. *Hearing Research*, *247*(1), 27–33. doi:10.1016/j.heares.2008.09.011

Worden, F., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42–52.

Zwicker, E. (1979). Different behaviour of quadratic and cubic difference tones. *Hearing Research*, *1*(4), 283–92.